

ISLS Annual Meeting 2026
June 15-19, 2026

Partnering with Purpose

19th International Conference on Computer-Supported Collaborative Learning (CSCL)

CSCL Proceedings

Edited by: Carolyn Rosé, Gahgene Gweon, Christa Asterhan, Anna Keune

ISLS 2026
June 15-19, 2026
Irvine, California

Supported by
UC Irvine School of Education
Wallace



International Society of
the Learning Sciences

International Society of the Learning Sciences

ISLS Annual Meeting 2026

Partnering with Purpose

Irvine, California, June 15-19, 2026
University of California, Irvine (UCI)

19th International Conference on Computer-Supported Collaborative Learning (CSCL) 2026

- CSCL Proceedings -

6th Annual Meeting
of the
International Society of the Learning Sciences (ISLS)

Editors:
Carolyn Rosé, Gahgene Gweon, Christa Asterhan & Anna Keune

International Society of the Learning Sciences

ISLS Annual Meeting 2026
Partnering with Purpose
Irvine, California, June 15-19, 2026
University of California, Irvine (UCI)

19th International Conference on Computer-Supported Collaborative Learning (CSCL) 2026

© 2026 International Society of the Learning Sciences, Inc. [ISLS]. Rights reserved.



www.isls.org

ISBN: 979-8-9906980-6-2 (CSCL Proceedings, PDF Version)
ISSN: 2543-0157

Cite as: Rosé, C., Gweon, G., Asterhan, C., & Keune, A. (Eds.). (2026). *Proceedings of the 19th International Conference on Computer-Supported Collaborative Learning - CSCL 2026.* Irvine, California, USA: International Society of the Learning Sciences.

All rights reserved. Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear the copyright notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior written permission of the International Society of the Learning Sciences. The International Society of the Learning Sciences is not responsible for the use which might be made of the information contained in this book.

Conference logo design by Nina Wishnok, Founder, Creative Director, N of Many.

International Society of the Learning Sciences

Conference Organizers and Committees

Local Organizing Committee

Kylie Peppler, *University of California, Irvine, USA*
Mimi Ito, *University of California, Irvine, USA*
Katie Salen Tekinbaş, *University of California, Irvine, USA*
Maggie Dahn, *University of California, Irvine, USA*

LOC Hybrid Committee

Shiyan Jiang, *University of Pennsylvania, USA*
Debbie Fields, *Utah State University, USA*
Jamie Gravell, *Chapman University, USA*
Rishi Krishnamoorthy, *University of Toronto, Canada*
Charles Logan, *Northwestern University, USA*
Peter Liu, *University of California, Irvine, USA*
Karis Jones, *Baylor University, USA*
Teo Keifert, *University of North Texas, USA, chair*

Essential Support

Nicole Hoover, *International Society of the Learning Sciences, USA, ISLS Conference Coordinator*
Alicja Żenczykowska, *Conference System Architect*
Paulo Blikstein, *Teachers College, Columbia University, USA, Conference System Chair*
Thomas Hillman, *University of Gothenburg, Sweden, EasyChair Guru*
Julie Herrick, *Connected Learning Lab UCI, USA*

Volunteer Committee Co-Chairs

Santiago Ojeda-Ramirez, *University of California, Irvine, USA*
Christina Kimmerling, *University of California, Irvine, USA*
Jasmine Tran, *University of California, Irvine, USA*

Repository Manager

Joel Wiebe, *University of Toronto, Canada*

Advisory Committee:

Pirita Seitamaa-Hakkarainen, *University of Helsinki, Finland*
Chris Hoadley, *University of Buffalo, USA*
Paulo Blikstein, *Teachers College, Columbia University, USA*
Ben Shapiro, *University of Washington, USA*
Stephanie Teasley, *University of Michigan, USA*
Liz Charles, *Dawson College, Canada*

Proceedings Copyeditors

Shiri Kashi, *University of Haifa, Israel*
Adiel Tammam Benjamin, *University of Haifa, Israel*

 © ISLS

International Society of the Learning Sciences

Senior Reviewers

Alwyn Vwen Yen Lee
Alyssa Wise
Anders Mørch
Betsy DiSalvo
Bram De Wever
Calixto Gutierrez-Braojos
Camillia Matuk
Carolyn Rose
Charlott Sellberg
Chris Teplovs
Christa Asterhan
Christopher Proctor
Claire Polo
Claudio Alvarez
Crina Damsa
Cynthia D'Angelo
Dalila Dragnić-Cindrić
Daner Sun
Davinia Hernandez-Leo
Eleni Kyza
Elizabeth Koh
Engin Bumbacher
Frank Fischer
Freydis Vogel
Gahgene Gweon
Gaoxia Zhu
Gautam Biswas
Gerry Stahl
Grace Lin
Hanni Muukkonen
Hans Christian Arnseth
Hiroki Oura
Hironori Egi
Hongzhi Yang
Ingo Kollar
Iris Tabak
Jeremy Roschelle
Jessica Andrews-Todd
Jim Slotta
Jody Clarke-Midura
Johanna Pöysä-Tarhonen
Jun Oshima
Jürgen Buder
Karen Huffman
Kathryn McCarthy
Kati Mäkitalo
Kyu Yon Lim
Longkai Wu
Magnus Hontvedt
Marcela Borge
Marcelo Worsley
Maria Beatrice Ligorio
Martin Greisel
Masataka Kaneko
Melissa Gresalfi
Michael Rook
Michelle Hudson
Michelle Wilkerson
Min Kyu Kim
Moegi Saito
Nikol Rummel
Ornit Sagy
Oskar Lindwall
Pantelis M. Papadopoulos
Philip Bell
Prajakt Pande
Preeti Raman
Prerna Ravi
Ritsuko Oshima
Robert Huang
Roberto Martinez-Maldonado
Rogers Kaliisa
Rolf Steier
Sahana Murthy
Sebastian Simon
Sebastian Strauß
Sébastien George
Sten Ludvigsen
Susan A. Yoon
Susan Bridges
Tanmay Sinha
Thérèse Laferrière
Toshio Mochizuki
Tugce Aldemir
Ulrike Cress
Xiaoqing GU
Yael Kali
Yannis Dimitriadis
Yu Xia
Yugo Hayashi

 © ISLS

International Society of the Learning Sciences

Reviewers

Alberto Montresor
Alejandra J. Magana
Alejandra Martínez-Monés
Alessandro Pagano
Alex Acquah
Alex Kumi Yebooah
Ali Keramati
Alina Sheppard-Bujtor
Allison Hall
Aloysius Ong
Alyshai Nadeem
Amine Hatun Ataş
Amy J. Ko
André Matutat
Andria Gentry
Anil Mammen
Anna Keune
Anthony Botelho
Armanda Lewis
Arooj Zaidi
Ashwin Tudur Sadashiva
Ayano Ohsaki
Azusa Nakata
Bahar Shahrokhian
Barbara Hall
Blessed Mhungu
Bunichi Otaki
Burcu Şener
Caiqian Zheng
Caitlin Morris
Carmen Zahn
Carolyn Rose
Casper Harteveld
Chaewon Kim
Changzhao Wang
Charlotte Beal
Cheng Tang
Cheng-Huan Chen
Chenglu Li
Chew Lee Teo
Chih-Hsuan Chang
Chih-Pu Dai
Chris Steadman
Christa Asterhan
Christina Stiso
Christoph Richter
Clausell Mathis
Corey Schimpf
Cristina Villa-Torrano
Daeun Hong
Daniel Wendel
Danielle Harlow
Darren Gergle
Davis Velarde-Camaqui
Dengkang Chen
Destiny Destiny Nnadozie
Diego Zapata-Rivera
Divanshi Priyadarshni Wangoo
Dor Abrahamson
Dorentina Dedushaj
Eduard Arnau
Efrat Ayalon
Elham Tajik
Elif Polat
Elli Papageorgiou
Eman Elbealy
Emily Kerzabi
Emma Anderson
Emmanuel Sepúlveda Guzmán
Emre Dinç
Enkelejda Kasneci
Enruo Guo
Erfan Farhadi
Evan Patton
Fan Chen
Fang Wen
Fanjie Li
Fengjiao Tu
Gennie Mansi
Gi Woong Choi
Gina Tesoriero
Gitit Indra Putra Wacana
Gregory McGowin
Guang Yang
Guotsai Hung
Ha Nguyen
Hae June KIM
Haihua Chen
Haiyang Xin
Hakeoung Hannah Lee
Hamideh Talafian
Hao-Chuan Wang
Haoming Wang
Hava Ben Horin Abramsky
Hawra Rabaan
Heather Leary
Hecong Wang
Hedieh Najafi
Heliya Ahmadi
Henrik Tjønn
Hideki Kondo
Ho Sue Kwon
Hongming Li
Hui-Chun Hung
Ibrahim Adisa
Idris Solola
Iman Mohammadi
Inés Alvarez Icaza Longoria
Injila Rasul
Insook Han
Ishika Ishika
Jaclyn Ocumpaugh
Jacob Davidsen
Jaesung Hur
Jaeyoon Choi
Jane Awuah
Jason Harron
Jason Reitman
Jaycee Sansom
Jeffrey Bush Tayne
Jennifer Scianna
Jeongwon Lee
Jeremy Bernier
Jessica Owens
Jessica Vandenberg
Jiamin Tang
Jiayan Zhu
Jie Cao
Jie Xu
Jinhee Kim
Jionghao Lin
João Adriano Freitas
Joey Huang
John Ajamobe
John López
John Ottina
Jongchan Park Park
Jorge Membrillo-Hernandez
Joseph Michaelis
Junior Anthony Bennett
Kaiju Kangas
Karla Trujillo
Karsten D. Wolf
Katelyn Suchyta
Kati Stoddard
Kaylah Facey
Kexin Li
Koichi Yasutake
Kyungri Kim
Laila El-Hamamsy
Larishia Johnson
Laura Romiti
Lei Ding
Leqi Li
Li Zhao
Li-Jen Wang
Lin Yan
Lingqi Tan
Linor Hadar
Linyuan Wang

 © ISLS

Liru Hu
Litong Zeng
Liu Dong
Lu Ding
Luis Montesinos
Luise Mehner
Mahsa Rahmati Masouleh
Mansurbek Kushnazarov
Mariama Henry
Matthew Duvall
Matti Tedre
Maya Persico
Mengxiao Zhu
Mengze Hong
Mia Čarapina
Michael Schneider
Michelle Lui
Miroslav Suzara
Mohammed Saqr
Mohd Anas Wajid
Morgane Chevalier
Moshi Wang
Myunghwa Jin
Na'ama Av-Shalom
Namrata Srivastava
Natalie Spence
Nga Hoang
Nitasha Mathayas
Noora Laakso
Nozomi Fujikawa
Olayinka Oyewole
Oleksii Voronkin
Omar Israel Mendoza Rosas
Osnat Atias
Palmyre Pierroux
Paras Sharma
Patrick Bolger
Paul Flynn
Paulina Haduong
Pauline Aguinalde
Pei-Yi Lin
Pengjin Wang
Pipob Puthipiroj
Pirita Seitamaa-Hakkarainen
Polly Lai
Qilin Wang
Qiuchen Yu
Qizhi Xu
Rabia Ibtasar
Rajeev Joshi
Ram Das Rai
Rawad Chaker
Reshma Balakrishnan
Rita Detrick
Robin Lange-Fazio
Rodrigo Hidalgo Amat Y Leon
Rossana Mantaras
Sachika Shibukawa
Sadra Jafari Ghalehkohneh
Saki Milton
Sara Torre
Sarah Jung
Sarah Lee
Sarita Singh
Sebastian Tews
Seik Oh
Selcuk Kilinc
Seth Van Doren
Shan Zhang
Shatha N. Alkhasawneh
Shengquan Yu
Shigen Shimojo
Shihui Feng
Shiyan Pang
Shiyao Wei
Shotaro Naganuma
Shruti Jain
Shuang Geng
Shuang Li
Shufang Tan
Shulong Yan
Siqi Shen
Siyu Long
Sonsoles López-Pernas
Sotheara Veng
Sreecharan Sankaranarayanan
Stefano Cacciamani
Stefanus Relmasira
Suprabha Jadhav
Susan Harrington
Tamara Tate
Taryn Williams
Teresa Cerratto Pargman
Tianyu Ma
Ti-Cheng Tsai
Tolulope Famaye
Tomislav Jagušt
Tong Li
Toni Earle-Randell
Valentina Nachtigall
Vanessa Echeverria
Vishwas Badhe
Vitaliy Popov
Wei Jia
Won Hee Lee
Xiang Li
Xiao Hu
Xiaojie Zhang
Xiaomeng Huang
Xiaotian Zou
Xinran Zhu
Xintian Tu
Xinyi Li
Xinyu Guo
Yanqiu Li
Yeonji Jung
Yicheng Huang
Yifan Zhang
Ying-Tien Wu
Yoojin Bae
Yoshiaki Matsuzawa
Younghoan Cho
Yuen Man Tang
Yunjeong Chang
Yutong Tan
Yuxin Chen
Yuzhe Zhou
Zhenyao Cai
Zhenyu Cai
Zifeng Liu
Zongran Wang

 © ISLS

Preface

This volume contains the proceedings from the long papers, short papers, posters, and symposia presented at the 19th International Conference on Computer-Supported Collaborative Learning (CSCL). This event was part of the Sixth Annual Meeting of the International Society for the Learning Sciences (ISLS2026), held in Irvine, California, USA. At this event, scholars and practitioners in the field of CSCL convened in a hybrid format to explore cutting-edge work, including research and practice, anchored in this year's conference theme of *Partnering with a Purpose*.

The 2026 ISLS Annual Meeting theme, "*Partnering* with *Purpose,"* resonates strongly with the field of Computer-Supported Collaborative Learning (CSCL), where learning is understood as a socially and technologically mediated process. The theme highlights how meaningful collaboration among learners, educators, researchers, communities, and intelligent technologies can support learning across diverse contexts. Within CSCL, these partnerships are central to designing learning environments that foster dialogue, coordination, and shared meaning-making while addressing issues of equity, participation, and collective agency. The theme also foregrounds the role of collaborative technologies in structuring and supporting these interactions across classrooms, workplaces, and online learning communities. And finally, as we consider a partnership between humans and AI agents, we deeply consider the complementarity between humans and AI, and get in touch with who we are uniquely as humans on a changing landscape. This final interpretation of the theme will be explored in the CSCL invited symposium entitled "*Reimagining the Disrupted Landscape in Light of Human-AI Complementarity* ".

The CSCL Proceedings feature long papers, short papers, posters, and symposia, all of which undergo rigorous double-blind peer review. We received a total of **177** submissions from **28** countries across Europe, Asia-Pacific, and the Americas. These submissions showcased a diverse array of research, design, and technological innovations in the field of CSCL. In terms of acceptance rates, **32**% of long paper submissions (**30** out of **93**), **28**% of short paper submissions (**17** out of **61**), and **70**% of poster submissions (**16** out of **23**) were accepted in their respective categories. Additionally, **4** symposia were accepted: **3** focused on CSCL and **1** hybrid symposium, in addition to **1** invited symposium. As a result, the CSCL Proceedings features **30** long papers, **52** short papers, **77** posters, **4** symposia, and one hybrid symposium.

To advance CSCL theories, technologies, and designs, the CSCL program brought together scholarship from education, computer science (including artificial intelligence [AI], learning analytics, and data mining), psychology, linguistics, anthropology, sociology, and more. The program featured contributions that focused on various topics, among them:

A total of 30 long papers were arranged into 7 themes. As Human-AI Collaboration is an aspect of the *Partnership* theme that is a vital area for CSCL today, two long paper sessions focused on Human-AI Collaboration, making it the most popular theme. As design of CSCL experiences is a key activity where *Partnership* is observed, one session focused on Design. As we explore both the genesis of CSCL experiences as well as the experiences themselves, we often conceptualize processes within networks, and thus one session focused on Networks. Three sessions focused on components of collaborative learning experiences, including Knowledge, Socio-Emotional Processes, and Peer Feedback and Assessment.

A total of 52 short papers were arranged into 11 themes. As short paper contributions are more focused than long paper contributions, these themes were often correspondingly more targeted and focused than their related long paper themes. Three separate short paper themes related to the broader theme of Human-AI collaboration, namely AI-Human Interaction, the Role of AI in CSCL, and Perceptions of Innovation and Change. Specific CSCL experiences included Collaborative Problem Solving and Collaborative STEM Inquiry. Sessions that focused on components of collaborative experiences included Collaborative Learning Processes, Group Dynamics and Processes, and Support for Collaborative Learning. To complement the long paper session on Design, we included a short paper session on Innovative Design for Learning. Some forward looking methods-focused contributions were presented in a Collaborative Learning Research Methods session.

Both Long and Short paper contributions for Collaborative Problem Solving were featured in a single session including two long papers and two short papers.

 © ISLS

A total of 76 posters were grouped into three complementary themes. *AI-supported learning and pedagogy* focused on how GenAI/agents directly scaffold learning and teaching, while *Collaboration analytics and orchestratio***n** covered methods and tools for measuring and coordinating collaborative activity. Finally, *Inclusive, cultural, and embodied learning design* highlighted equity- and identity-attentive designs where culture, affect, and embodiment shape participation and meaning-making.

As **keynote** speaker for CSCL 2026, Alyssa Friend Wise reflected the conference's commitment to advancing purposeful partnerships in technology-rich learning. A leading scholar whose work spans CSCL, learning sciences, data science, and human-centered design research, Professor Wise represented the theme of "Partnering with a Purpose" through her extensive experience fostering meaningful collaborations across disciplines, institutions, and stakeholder groups. Her keynote highlighted how innovation in CSCL emerges not from isolated efforts, but from diverse forms of partnership, including institutional collaborations, educator partnerships, and interdisciplinary work. Through the articulation of three kinds of partnerships (i.e., Flexible Coupling, Constructive Responsiveness, and Generative Pluralism), Professor Wise provided a conceptual lens and practical guidance for understanding how collaborative relationships across multiple dimensions of the research process can drive impactful CSCL research. Her boundary-spanning perspective inspired exploration regarding how purposeful partnerships can shape CSCL research.

As mentioned above, *Reimagining the Disrupted Landscape in Light of Human–AI Complementarity* was the title of the **invited CSCL 2026 symposium**. Facilitated by Carolyn Rose (CMU / USA), the invited speakers, Professor Susan Jurow (UC Boulder / USA), Professor Shayan Doroudi (UCI / USA), and Wenli Chen (NIE / Singapore), discussed the evolving role of artificial intelligence in CSCL. The symposium positioned felt collaborations with AI as a generative opportunity to rethink learning and collaboration. Centered on questions of how human and AI capacities can complement rather than compete, the symposium explored models of hybrid collaboration that foregrounded human judgment, creativity, and agency while leveraging AI. It further examined how learning environments and educational policies might be redesigned to support equitable and trustworthy human-AI learning. Drawing on empirical research, design cases, and foundational theoretical perspectives of learning, the symposium presented how technological disruption can become a catalyst for resilient, participatory, and humane forms of collaborative learning in partnership with AI.

We would like to express our sincere gratitude to all the individuals who dedicated many hours to ensure the high quality of the program. We appreciate the efforts of the **291** reviewers and **90** meta-reviewers who conducted over 662 reviews and meta-reviews for CSCL. Their thorough and constructive feedback was essential to the program's success. We would also like to thank the hybrid engagement subcommittee, chaired by D. Teo Keifert, for their invaluable support in reviewing the CSCL hybrid symposium.

We wholeheartedly thank the ISLS, especially Nicole Hoover, for their guidance, as well as that of the ISLS committees, especially the CSCL Community Committee. Special thanks also go to Alicja Żenczykowska and Paulo Blikstein for the important continuing work towards revamping the submission and review process. And of course, none of this would have been possible without the tireless work of the UC Irvine local organizing team. Zot zot, in the industrious and joyful UC Irvine spirit!

Carolyn Rose, Christa Asterhan, Gahgene Gweon, and Anna Keune
Program Co-chairs, CSCL 2026

 © ISLS

Acknowledgments

ISLS Annual Meeting 2026 conference organizers would like to give special thanks to the following committees and people for their support:

- International Society of the Learning Sciences (ISLS)
- ISLS Board of Directors and Executive Committee
- University of California, Irvine
    - School of Education
    - Department of Informatics
    - Connected Learning Lab
- Conference Organizing Committee including:
    - ICLS and CSCL Program Committees
    - ISLS Hybrid Engagement Sub-Committee
    - Volunteer Committee Co-Chairs
    - Doctoral Consortium Co-Chairs
    - Early Career Workshop Co-Chairs
    - Mid-Career Workshop Co-Chairs
    - Alt-Academia Workshop Co-Chairs
    - Interactive Tools and Demos Co-Chairs
    - Pre-Conference Workshops and Tutorials Co-Chairs
    - Arts & Gallery Co-Chairs
    - ISLS 2026 Essential Support Team
- Reviewers and Meta Reviewers
- Outstanding Paper Award Committee
- Conference Proceedings Copy Editors
- Wallace Foundation
- Gordon and Betty Moore Foundation
- National Science Foundation

 © ISLS

International Society of the Learning Sciences

Table of Contents

CSCL Proceedings

Special Session

CSCL Invited Symposium 3
A. Susan Jurow, Wenli Chen, Shayan Doroudi, Carolyn Rose

Long Papers

Designing Multi-Agent AI Systems for Collaborative Learning: Strategic Questioning Patterns and Cognitive Development in Large-Scale Online Learning 6
Xiaojie Niu, Jingjing Zhang

Structural Signals and Role Archetypes: The Emergence of Roles in a Distributed Epistemic Game 15
Jenna Matthews

Biased Chatbot Interaction as Preparation for Collaborative Learning 23
Tanmay Sinha, Kumaran Rajaram

How Simulation Design Shapes Academically Productive Talk: An Epistemic Network Analysis of Teacher Discourse in Middle-School Physics 32
Wei Jia, Kexin Li, Gaowei Chen

Supporting Students' Shared Epistemic Agency Discourse with GAI via Learning Analytics Augmented Meta-discourse 42
Yawen Yu, Can Sun, Gaowei Chen

Collaboration and Competition in Educational Gameplay: Strategic Student Decision-Making 51
Tianyu Ma, Jennifer Kahn

Exploring Variations in Teacher Modelling and Scaffolding Through Chatbot Interactions in Middle School Classrooms 61
Henrik Tjønn, Sten Ludvigsen, Anders Mørch

Peer Feedback's Role in Undergraduates' Interdisciplinary Collaborative Learning Outcomes and Regulated Learning Patterns 70
Guo Su, Wenli Chen, Gaoxia Zhu, Yubei Shi, Li Wei, Chencheng Le

How Do Pre-Service Teachers Integrate Multiple Peer Feedback Messages? Identifying Different Integration Patterns 80
Julia Hornstein, Martin Greisel, Ingo Kollar

Student, Hybrid, or AI?: Examining Authorial Roles and Network Structures in GenAI-Supported Collaborative Learning 90
Shaoming Chai, Zhenhai He, Tongwu Jiang

Towards Teachers-AI Collaboration for Professional Development in Learning Design 99
Shatha N. Alkhasawneh, Davinia Hernández-Leo

Socially Shared Metacognitive Regulation (SSMR): Comparative Analysis of Strategies and Adaptations in Teams with Divergent Outcomes during CPS 109
Vishwas Badhe, Ramkumar Rajendran

Close Encounters: Teacher Proximity and Student-Reported Emotions While Using ST Math 119
Sotheara Veng, Yiqin Chao, Rosa Mykyta-Chomsky, Teomara Rutherford

Farm Harvests and Airplane Engines: Contextualizing Database Exercises in Relatable Contexts with a Learnersourcing Workflow 129
Yuzhe Zhou, Alejandra J Magana, Tianyi Li

Bridging Curriculum and Creativity: NLP-Powered Semantic Analysis for Knowledge Building 139
Ahmad Khanlari

 © ISLS

An Item Response Theory Model for Multiple Resubmissions 148
Xiang Liu, Mo Zhang, Chen Li, Hongwen Guo, Amy Ko, Min Li

Estimating Teacher Orchestration Load in Co-located and Online CSCL: Multimodal Data Analysis 155
Lubna Hakami, Ishari Amarasinghe, Davinia Hernández-Leo

Bridging Group and Individual: Transitional Spaces in Collaboration 165
Xuesong Cang, Michael Tissenbaum, Sadhana Puntambekar

When Peers Outperform AI (and When They Don't): Interaction Quality Over Modality 175
Caitlin Morris, Pattie Maes

Frame Alignment and Feedback in CT-ified Science Classrooms 185
Connor Bain, Uri Wilensky

A Multimodal Study of Emotion Regulation in Student Dyads During Computational Model Building 194
Ashwin T S, Srigowri Mayasandra Prasanna, Joyce Fonteles, Gautam Biswas

Exploring Cognitive, Social, and Metacognitive Dimensions of Student-Student-LLM Collaboration in Problem Solving 204
Victoria Delaney, Ibrahim Oluwajoba Adisa

Quantifying Knowledge Convergence through Change Point Detection of Word Abstractness in Collaborative Discussions 214
Ryunosuke Nishimura, Hironori Egi

Small Arguments, Big Ideas: Fostering Children's Collaborative Reasoning with LLM Agents 224
Hecong Wang, Meng Wang, Erfan Farhadi, Asad Shahab, Yuanzhu Li, Haozheng Du, Carolyn Rose, Hangfeng He, Zhen Bai

A Design Space for Dialogic Assessment in CSCL 234
Rolf Steier, Magnus Hontvedt, Jacob Davidsen, Charlotte Beal, Morten Oddvik, Marit Solem

Flexibility Before Fluency: Exploring How Flexible Strategies Emerge through Collaborative Problem-Solving 244
Seiyon Lee, Shan Zhang, Hongming Li, Ji-Eun Lee, Avery Closser, Anthony F. Botelho

Using Multimodal Analytics-Supported Reflective Assessment to Promote Undergraduates' Productive Social-emotional Interaction 254
Yuqin Yang, Xiaoru Li, Daner Sun, Xueqi Feng, Gaoxia Zhu

Effects and Mechanisms of Learning Analytic-Supported Reflective Assessment for Promoting Metacognitive Regulation among Undergraduates 264
Yuqin Yang, Yangbo Zhao, Xueqi Feng, Daner Sun, Jianwen Sun

Extending CSCL Affordances for Multilingual Classrooms: Teacher Perspectives on AI-Supported Collaboration 274
Pragati Maheshwary, Alina Guha, Shamya Karumbaiah

Integrating Augmented Reality and Product Design in Makerspaces 283
Dejan Trencevski, Jina Kang, Vishal Sachdev, Aric Rindfleisch

Short Papers

Unpacking Flexibility in Teacher Professional Learning on WhatsApp Messaging Network Platform 294
Iris Saban, Linor Hadar

Teachers' Evolving Sentiments and Competencies about Generative AI in Social Media Discourse: A Case Study on X 299
Karina Vázquez, Davinia Hernández-Leo, Shatha N. Alkhasawneh

Advancing Dialogic Leadership through AI-Enabled Training: Insights from a Design-Based Research Project 304
Xuanning Chen, Imogen Casebourne, Rupert Wegerif, Pedro Salinas

Cultivating Boundary-Crossing STEM Competencies: Building Community and Embracing Other

 © ISLS

Ways of Knowing in Global Informal Learning 309
Danielle P. Espino, Kristina M. Lux, Seung B. Lee, Eric R. Hamilton

Age Differences in Cognitive Engagement and Their Association with Learning Performance in Collaborative Learning with Concept Maps: Experimental Investigation Using Epistemic Network Analysis 314
Shigen Shimojo, Yugo Hayashi

From Idea Consumers to Knowledge Co-Creators: A 3-Year Design Study of Sustained Creative Work Through Cross-Community Partnership 319
Robert Huang

Reconceptualize Conversational AI as a Learning Medium, From a Holistic Learning Perspective 324
Yuqi Wang

Scoping Review of Knowledge-Creation Practices Supported by AI: Identification of Research Strands of Different Communities 329
Jun Oshima, Shotaro Naganuma

Self-organizing Interdisciplinary Team in Generative AI-Supported Collaboration Context 334
Shotaro Naganuma, Tsubasa Minematsu, Jun Oshima

Exploring Shared Agency in an Intelligent Immersive Learning Environment for Collaborative Writing 339
Yun Wen, Zongran Wang, Lang Xu

Student Perceptions of Generative AI Chatbots in Computing Education: A Technology Acceptance Lens .. 344
Ean Teng Khor, Elizabeth Koh, Leta Chan

Combining Collaborative and Adaptive Learning: Constructive–Interactive Engagement through Concept Maps, Teachable Agents, and an Adaptive Tutor 349
Yugo Hayashi, Shigen Shimojo, Kenneth R. Koedinger

When Ideas Take Turns: Temporal Embodiment of Disciplinary Engagement 354
Xiaotian Zou, Daeun Hong, Cindy E. Hmelo-Silver, Chen Feng, Krista Glazewski, Tianshu Wang, James Lester

Participatory Design for AI in Teacher Education: From Top-Down Directives to Teacher Empowerment ... 359
Ekta Shokeen, Ane Vielandt Jensen, Ole Sejer Iversen, Christian Dindler, Marianne Graves Petersen

DotQuest: Accessible STEM Games to Support Blind and Visually Impaired Learners 364
Ariel Han, Rajeev Joshi, Daniel Tsivkovski, Dylan Ravel, Maryam Etezadbrojerdi, Franceli Cibrian

Pieces of the Puzzle: Examining Multimodal Interactions in Two Jigsaw Tasks to Inform Collaboration Support 369
Indrani Dey, Collette Heskett, Mon-Lin Monica Ko

Reconceptualizing Activity Theory for Human-AI Teaming in Computer-Supported Collaborative Learning 374
Jaeyoon Choi, Nia Nixon

Designing Digital Enhancements for Social Emotional Skill in a Collaborative Learning Environment 379
Enruo Guo, Emily Kerzabi, Jessica Andrews-Todd, Jiangang Hao, Patrick Kyllonen, Stephen Elliott

Linking Physiological Data and Self-Reports: Exploring Engagement in Collaborative Learning 384
Xiaoran Li, Serkan Ucan

GenAI Chatbot for Socially Shared Regulation of Learning in CSCL: Impact on University Students' Learning Engagement and Outcome 389
Dengkang Chen, Yi Zhang, Jianwen Sun, Yijie Lu, Xiaoge Guo, Fucheng Guo

Regulating Epistemic Emotions in CSCL: Comparing the Higher- and Lower-Performing Groups 394
Wenli Chen, Lishan Zheng, Xuanyu Chen, Qianru Lyu

The Impact of Shared Feedback Visualization on Student Reflection During Simulation Debriefing 399
Vitaliy Popov, Anouschka van Leeuwen

 © ISLS

Exploring Unstructured Collaboration While Navigating Misinformation in Computer-Supported Collaborative Argumentation: Implications for Design 404
Ram Das Rai, Sahana Murthy

Multi-Level Knowledge Building Engagement: Pathways to Collaborative Problem-Solving and Academic Excellence in Higher Education 409
Chih Hui Seet, Chih-Hsuan Chang, Huang-Yao Hong

Examining Socio-Cognitive and Affective Dynamics of Team Functioning and Leadership in Medical Simulations 414
Annamarie Brosnihan, Megan Wiedbusch, Cameron Marano, Lindsey Schulz, Anna Pirner, Shiva Kalidindi, Roger Azevedo

Leveraging Generative AI for Iteration in Undergraduate Design 419
Inés Álvarez-Icaza, Claudia Susana López Cruz, Martha Elena Núñez López, Griselda Esthela Oyervides Ramírez, Juan Carlos Márquez Cañizares, Christiam Ivan Mendoza García

From Regulation to Co-Regulation: A Game-Based Simulation for Human and Simulated Learners during Complex Problem Solving 424
Roger Azevedo, Megan Wiedbusch, Cameron Marano, Annamarie Brosnihan, Milouni Patel

Knowledge Network for Innovations in Learning and Teaching (KNILT): Teacher Learning Through Knowledge Creation 429
Ying Kong, Jianwei Zhang, Aya Isaac

Learning Across Levels: Conceptualizing Transitional Space 434
Xuesong Cang, Sadhana Puntambekar

From Words to Networks: AI-Powered Knowledge Graphs for Understanding Collaborative Learning 439
Ahmad Khanlari

How Active User Count Relates to Group Outcomes in Collaborative Programming: An Implicit Scripts Perspective 444
Zhenyu Cai, Richard Lee Davis, Roland Tormey, Pierre Dillenbourg

From Initiative to Impact: Modeling How Collaborative Initiative Shapes Problem-Solving in STEM Simulations 449
Rodrigo Hidalgo, Jina Kang

A Transpositional Grammar Cyber-Social Study of Student–AI Relationships 454
Christopher Hughes, Akash K. Saini

Who Gets to Know? Children's Epistemic Agency in Science eBook Reading with AI 459
Xuechen Liu, Ying Xu, Jon M. Wargo

A Case Study on the Effect of TAs' Awareness of Pre-class Report Submissions on Scaffolding in Small-group Discussions 464
Nozomi Fujikawa, Sanii Mizuno, Shintaro Uchiyama, Yusuke Morita

Reframing Social Annotation as a Text-Embedded Collaboration Analytics Ecology: The Idea Creation Cycle in Social Annotation 469
Yeonji Jung, Sophia Soomin Lee

Embodied and Non-Embodied Technologies in Mathematics Learning: A Comparative Study 474
Injila Rasul, Ivon Arroyo

From Solo to Collaborative: Design Affordances in Multi-User Immersive VR Simulations for Transforming Science Learning 479
Michelle Lui, Sunnie Gong, Chloe Lok

Epistemic Authority and Interface: A Typological Framework of AI's Roles in Knowledge Building 484
Lydia Cao, Bodong Chen, Sebastian Simon, Yee Yin Tan, Stefano Cacciamani, Eliane Chaker Metni, Wakana Tsuji, Robert Huang, Seng Chee Tan, Caroline Traverse, Shotaro Naganuma, Jun Oshima, Thomas Underwood, Adiel Benjamin, Yotam Hod, Elaine Khoo, Thérèse Laferrière, Carl Bereiter, Marlene Scardamalia, Richard Reeve, Ahmad Khanlari, Boris Streipe

Modeling and Analyzing Collaborative Dialogue States in Mathematics Problem-Solving through ONA and Hidden Markov Models 489
Shan Zhang, Andrés Felipe Zambrano, Hongming Li, Seiyon Lee, Ji-Eun Lee, Anthony F. Botelho

Designing for Collaboration: An Engineering Notebook to Scaffold Collaborative Problem-Solving in Middle School Robotics 494
Toni Earle-Randell, Chris Palaguachi, Joanne Barrett, Jennie Lee, Jina Kang, Maya Israel, Emmanuel Dorley, Kristy Elizabeth Boyer, H. Chad Lane

"It Wasn't So Outside of the Box, But It Wasn't Something I Would Have Thought About": AI as a Creative Collaborator in Student Problem Solving 499
Ibrahim Oluwajoba Adisa, Victoria Delaney

Are We There Yet? The Roles and Development of GenAI-Based Tools Supporting Collaborative Learning: A Scoping Review 504
Maya Persico, Daeun Hong, Jerika Miller, Cindy E. Hmelo-Silver, Freydis Vogel

Collaboration Quality in Computational Modeling: Implications for Middle-School STEM+C Learning 509
Shruti Jain, Namrata Srivastava, Srigowri Mayasandra Prasanna, Clayton Cohn, Umesh Timalsina, Naveeduddin Mohammed, Gautam Biswas

Scaffolding Knowledge Building through Ordered-Epistemic Network Analysis-Supported Reflective Assessment and Meta-Discourse 514
Xiaojie Zhang, Carol K.K. Chan, Shihui Feng, Jianhua Zhao

Facilitating High School Students' Socially Shared Regulation of Learning Through STEM Inquiry into Local Air Quality and Asthma 519
Ho Sue Kwon, Catherine Roeroe, Susan A. Yoon

MIRACLE: Multi-Agent Intelligent Regulation to Advance Collaborative Learning Environment 524
Shuang Li, Haiyang Xin, Yimeng Sun, Qiannan Niu, Lingyun Huang, Gaowei Chen, Yibing Zhang, Ching Sing Chai

Unveiling Role Dynamics in AR-Based Collaborative Inquiry 529
Xinyue Jiao, Zifeng Liu

Inquiry Script Modeling: A Formalism for Scripting and Orchestration of Collective Inquiry 534
Erfane Ghasempour Khoshroodi, James D. Slotta

The Effects of GenAI-Assisted Programming Learning on Fifth Graders' Computational Thinking, Programming Self-Efficacy and Outcomes, and Programming-oriented Prompting: A Comparison of Two Human–AI Collaboration Configurations 539
Ying-Tien Wu, Yi-Hong Zeng, Ching-Hui Kuan

STEM Teachers' Perspectives on Generative AI: A Social Media Content Analysis of Reddit Discussions .. 544
Leqi Li, Xin Guan

Teacher Perception and Practices of the Next Generation Science Standards (NGSS): Insights from a New England Study 549
Anupam Raj, Michael Wittmann

Symposia

Designing AI Partner-Embedded Learning Environments to Support AI Literacies 556
Mon-Lin Monica Ko, Marcelo Worsley, Michael Alan Chang, Chelsea Chandler, Sierra Rose, Brooklyn Clines, Emily Watts, John Weatherley, Peter W. Foltz, Rohit Raju, Roshan Muddaluru, Jeffrey Bush Tayne, Gregory Benedis-Grab, Quentin Biddy, Thomas Breideband, William R. Penuel, Tamara Sumner, Collette Heskett, Nga Hoang, Julia Eden, Sidney D'Mello

Revisiting Wise and Schwarz's Provocations for CSCL: Partnerships for Transformation and Educational Change 566
Breneil Malcolm, Marcela Borge, Alyssa Wise, Marcelo Worsley, Sebastian Simon, Tanner Vea, Natalie Rae, Martin Greisel, Mia Čarapina, Ban Mouid Shiwalia

 © ISLS

Rise Above State-of-the-Art: Designing New Possibilities of GenAI for Knowledge Building 577
Lydia Cao, Dina Soliman, Carol K.K. Chan, Carl Bereiter, Marlene Scardamalia, Shotaro Naganuma, Tsubasa Minematsu, Jun Oshima, Jiayu Cheng, Vivian Leung, Yue Zhao, Chen Wang, Bodong Chen, Thérèse Laferrière, Stéphane Allaire, Adiel Benjamin, Yotam Hod, Jianwei Zhang, Yanqing Sun, Sudarshan Kammu, Mei-Hwa Chen, Thomas Underwood, Seng Chee Tan, Yee Yin Tan, Chew Lee Teo, Guangji Yuan, Xueqi Feng

Towards a Robust Theory Toolbox for CSCL: Mechanisms, Models, and AI Models, and AI 589
Sten Ludvigsen, Davinia Hernández-Leo, Peter Reimann, Alyssa Wise, Fanjie Li, Hans C. Arnseth, Kenneth Silseth, Rolf Steier, Jo Inge J. Froytlog, Michael Baker

Posters

Partnering Through Ceremony: A Multimodal Design for Collaborative Healing and Learning 603
Karla Trujillo

Using Artificial Intelligence as a Critical Thinking Partner: A Conceptual Demonstration 606
Spenser Clark

Trans-Ecological Disruptions and Pedagogical Interventions in GAI-Supported Collaborative Learning 609
Yue Han, Wen Qi

Fostering Cocreativity in Primary Education Through Viacocrea: Effects on Students' Creative Products 612
Maria Sitjà Martín, Manoli Pifarré Turmo

AI as a Collaborative Partner: EMPA's Role in Supporting Reflective Teamwork Learning 615
Aparajita Jaiswal, Ashish Ashish

Boundary Spanning Narratives in a Constructionist Museum Game 618
Eda Zhang, Matthew Berland, Yunseo Lee

Neurodiversity in Computer-Supported Collaborative Learning: A Literature Review 621
Greta Heathcote, Alison Nelson, Armando Paulino Preciado Babb

What Draws Upvotes? Examining Relationships between Interpersonal Skills and Social Annotation Behaviors in Online Settings 624
Hanxiang Du, Jiyoung Cheong, Gaoxia Zhu, Xiaoshan Huang, Shan Li, Juan Zheng, Jacqueline Tan, Tianlong Zhong, Chenyu Hou, Tongxi Liu

Exploring STEM Teachers' AI Readiness, Attitudes, and Uses of Generative AI across Teaching Subjects 626
Lin Yung-Hsuan, Lin Pei-Yi

The Relationship between Argumentative Discourse Sequences with Generative AI and Students' Argumentative Writing Structure 629
Kyungri Kim

Belonging Infrastructure: A Designable Construct for Cultivating Teacher Agency in RPPs 632
Gregory Benedis-Grab, Alexandra Gendreau Chakarov

Resolving Conceptual Contradictions through Transformative Integration 635
Weiyi Xiao

Rotating Leadership of Transactive Discourse through Synchrony and Desynchrony of Multimodal Interaction 638
Jun Lu, Jun Oshima, Ritsuko Oshima

Dialogic Intelligence: Rethinking the pedagogical goal of CSCL in the context of Generative AI 641
Rupert Wegerif, Xuanning Chen, Bo Yu, Imogen Casebourne, Kevin Martin

Designing for Augmenting Teachers' Understanding of Children with ADHD Behaviors Using a Smartwatch 644
Lauren Min, Ariel Han, Katelyn Teav, Daisy Fernandez-Reyes, Franceli Cibrian

 © ISLS

"He's Gonna Say It's Trash": Students' Emotional Experiences with AI Feedback in Collaborative Scientific Inquiry 647
Megan Humburg, Namrata Srivastava, Sarah Burriss, Yeo Jin Kim, Joshua Danish, Cindy E. Hmelo-Silver, Shruti Jain, Clayton Cohn, Umesh Timalsina, Krista Glazewski, Gautam Biswas, James Lester

Iterative Design for Distributed Embodied Collaboration 650
Casey Smith, Mike Tissenbaum

Complementary Roles of Collaborative Reflection and AI-generated Feedback to Support Students' Science Writing 653
Indrani Dey, ChanMin Kim, Sadhana Puntambekar

Collaborative Virtual Reality Worlds, Cultural Play: Re-imagining Game Design through Interactional Tool Exploration 656
Xingxing Xie, Heliya Ahmadi, Dylan Paré

At the Intersection of Cognitive Engagement and Critical Thinking in AI-Assisted Learning 659
Lu Ding, Erika Cheek, Tami May

ClassMosaic: Leveraging Solution Diversity to Teach Computational Problem-Solving 662
Laila El-Hamamsy, Engin Bumbacher

Multimodal Analysis of Mathematical Statistics Learning in a CSCL Environment 665
Masataka Kaneko, Hironori Egi, Takuya Kitamoto, Takeo Noda

Leveraging CSCL for Social Emotional Learning: Early Insights from a Large-Scale Classroom Study 668
Jiangang Hao, Patrick Kyllonen, Emily Kerzabi, Wenju Cui, Enruo Guo, Jessica Andrews-Todd, Yang Jiang, Stephen Elliott

Collaborative High School Teacher Learning in Bichronous Online STEM Professional Development 671
Ho Sue Kwon, Susan A. Yoon, Jooeun Shim

Enhancing K-12 Teachers' Confidence and Competence in STEM Integration Using Blended Learning in Professional Development 674
Haowei Li, Amanda M. Cottone, Susan A. Yoon, Zheng Bian, Rebecca G. Wells

Knowledge-Building Discourse in Generative AI-Mediated Collaborative Problem Solving: A Student–Expert Comparison 677
Bunichi Otaki, Ritsuko Oshima, Jun Oshima

Integrating AI and Interaction Analysis: A Mixed-Methods Approach to Tracing Object-Based Participation in CSCL 680
Hakeoung Hannah Lee, Miguel Lujan, Venkatesh Jatla, Ugesh Egala, Marios Pattichis, Sylvia Celedon-Pattichis

Human–AI Learning Alignment in Situated Contexts: A Human-Centered Framework 683
Shi Ding, Brian Magerko

Calling, Collecting, and Transformation: Playful Artifacts as Pivots for Joint Activity and Game Design in a Collaborative VR Design Studio 686
Heliya Ahmadi, Xingxing Xie, Dylan Paré

Investigating support for collaboration on the AI Frontier: A Qualitative Study of Middle-Grades Computer Science Classrooms Informed by the MOSAIC Protocol 689
Jeffrey Bush Tayne, Nga Hoang, Indrani Dey

Physiological Indicators of Teacher Stress in CSCL Versus Non-Collaborative Learning Sessions 692
Eduard Arnau Romeu, Batuhan Sayis, Davinia Hernández-Leo

Beyond Teacher-AI Dyads: A Triadic Synergy Model for Transforming Teacher Professional Development 695
Kexin Li, Gaowei Chen

 © ISLS

The Effect of Group-Awareness on Peer Feedback Quality in Pre-Service Teachers' Collaborative Diagnostic Reasoning 698
Sebastian Tews, Melissa Özsoy, Meral Roeben, Yasemin Erdemgil, Xiao Zhang, Christiane Lütge, Freydis Vogel

Comparing High School Students' Collaborative Discourse in Idea-Centered and Procedure-Based Technology-Supported Scientific PBL 701
Ying-Tien Wu, Ti-Cheng Tsai, Li-Jen Wang

EnactingBoard: AI-mediated Dialogue for Developing Empathic Negotiation Skills in Higher Education 704
Hideyuki Suzuki, Hideo Funaoi, Yoshihiko Kubota, Hiroshi Kato

Fostering Metacognitive Awareness for Effective Human–AI Collaboration in Learning 707
Pau Benazet i Montobbio, Davinia Hernández-Leo

Hierarchical Multi-Agent System for Instructional Design in Music Knowledge Building 710
Menglian Zhang, Mingyang Zhu, Haiyang Xin, Yibing Zhang

AI Tools in Arts Learning: A Scoping Review 713
Jaycee Sansom, Safinah Ali

Principle-Driven Scaffolding: Implementing and Evaluating LLM-Generated Micro-Scripts in an Informal Learning Community 716
Hideki Kondo

Equitable Learning Analytics Dashboards: Roles and Mechanisms in Personalized Learning 719
Wei Yan, Hyejeong Lee, Donna Dancer, Wonjin Yu

Adaptive Support for Collaborative Learning: Effects of Design Features Across Task Types 722
Lixiang Gao, Frank Fischer, Olga Chernikova

Agentic Collaboration with GenAI: A Cautionary but Optimistic Tale 725
Iris Tabak, Gur Arie Livni Alcasid, Ilana Dubovi

Sociotechnical Efficacy in Modeling 3D Interactive Environments 728
Akin Adediji, Jinhee Lee

Designing an AI-Augmented Knowledge Synthesis Tool for Collaborative Learning 731
Xinran Zhu, Bo Shui, Ishita Asnani, Liam Magee

Encouraging Deeper Learning with GenAI: A Review of Pedagogical Constraints on GenAI and Their Impact in Higher Education 734
Shuang Geng, Di Fan, Juhong Eom, Gahyun Callie Sung

Technological Competency: A Conceptual Model for Understanding the World Around Us 737
Matthew Duvall, Chen Wang

Idea Tracing Analysis (ITA): A Temporal-Structural Methodology for Modeling Collaborative Discourse and Contextualizing Multimodal Interactions 740
Xiaomeng Huang, Xavier Ochoa, Madhumitha Gopalakrishnan, Shiyu Zhang

Do Anti-democratic Bots Dream of Digital Riots? Modeling Anti-democratic Decision-making and Collective Knowledge Construction Using Natural Language Processing 743
Renato Russo, Carolina Sotério, Fabio Campos, Paulo Blikstein, Gustavo Oliveira

Exploring Collaborative Teacher-AI Interaction for Rehearsing Responsive Teaching in Elementary Mathematics Education 746
Shuman Wang, Gordon Yeung, Miroslav Suzara, Jennifer Osuna, Charis Charitsis, John Mitchell

The Impact of Conversational Agents on Children's and Parents' Engagement During Shared Reading 749
Dorentina Dedushaj, Kathryn Leech, Isabella Ruel, Grace Lin

Novel Figurative Language to Identify Promising Ideas in Knowledge Building Scientific Discourse 752
Wakana Tsuji, Marlene Scardamalia, Ahmad Khanlari

Augmentation, Enhancement, or Replacement? Classifying Generative AI Use in Workforce Development Practice 755
Jacob Hale, Tessa Forshaw, Mari Longmire

Investigating Collaborative Weaving Design as a Scaffold for AI Literacy 758
Carolyn Rose, Olivia Robinson, Melisa Orta-Martinez, Joey Huang, Kylie Peppler

Reflective Assessment for Authentic Research among High School Students: Enhancing Epistemic Agency for Knowledge Building 761
Lei Ding, Jianwei Zhang, Thomas Underwood, Weiting Qi, Dongni Guo, Alae Sherafat, Lana Hower, Jennifer Muirhead, Maryann Dunn

Using Generative AI as an Epistemic Catalyst for Collaborative Lesson Planning 764
Yuxin Chen, Drew Gossen

Assessing Student Collaboration with Multimodal Data and Large Language Models 767
Shan Zhang, Hongming Li, Seiyon Lee, Ji-Eun Lee, Noah L. Schroeder, Anthony F. Botelho

A Teacher-in-the-Loop Multi-Agent Framework for Transparent Essay Scoring and Feedback 770
Ali Keramati, Mark Warschauer

Fostering Preservice Teachers' Design Thinking by Integrating Generative AI and Analytics-Supported Reflective Assessment in a Collaborative Inquiry Environment 773
Yuqin Yang, Linbaiyu Chen, Daner Sun, Wangda Zhu, Shiyan Pang

Shared Epistemic Agency in A Collaborative Game for AI Literacy 776
Yue Shi, Lingfei Cao, Zhichun Liu

Exploring Tangible Interfaces for Learning 3D Geometry and Basic Structural Engineering Concepts 779
Arooj Zaidi, Sandra Yu Okita, Giulia Barbareschi, Chihiro Sato, Junichi Yamaoka

Developing Postgraduates' Feedback Literacy through Multiple-sourced Feedback Practices in an L2 Writing Class 782
Chunlin LEI, Fuxin KONG

Introducing IDEAL: An Intelligent Design Environment for Active Learning 785
Chenxuan Meng, James D. Slotta

Analyzing Human–AI Interaction Patterns in Generative-AI-Supported Inquiry-Based Learning 788
Ayano Ohsaki, Hideya Matsukawa, Ryohei Egusa, Satoshi Takahashi

Between Modalities: Observing How Learners Move between Verbal and Digital Spaces in Face-to-Face Collaborative Knowledge Building 791
Litong Zeng, Yiqi Xiao, Mike Tissenbaum

One Size Fits All? Learner-Adaptive Peer Feedback and Perceived Feedback Adequacy in Pre-Service Teacher Education 794
Alina Sheppard-Bujtor, Martin Greisel, Julia Hornstein, Ingo Kollar

If One Hack Week is Pedagogically Effective, Why Not Ten? 797
Jacob Puthipiroj, Thepchai Supnithi, Nongnuch Ketui

Integrating Individualized Education Programs into AI-based Tutoring Systems: An Interview Study with Teachers 800
C. Naomie Williams, Margaret Ellen Seehorn, Haiyi Zhu, Vincent Aleven

Computational Modeling of the Process of Scientific Inquiries 803
Nanae Watanabe, Masahiro Tada, Shin'Ichi Konomi, Keisuke Koba, Masaya Okada

Orchestrating a Web-Based Classroom Game: Balancing Platform Automation and Teacher Work 806
Kai Yan, Jasmine Hu, James D. Slotta

Meet Bip: Understanding Student–Peer Agent Interaction in an Embodied Exploratory Learning Environment 809
Erfan Farhadi, Kenneth Fei, Yifan Jiang, Zhen Bai

 © ISLS

Dropout and Success in Online Higher Education: A Comparative Analysis of Students in Brazil 812
Andréia Maria de Lima Assunção, Paula Daniela Zapata, Lívia Pizauro Sanchez

Author index A-1

 © ISLS

International Society of the Learning Sciences

Special Session

 © ISLS

 © ISLS

International Society of the Learning Sciences

CSCL Invited Symposium

A. Susan Jurow, Wenli Chen, Shayan Doroudi, Carolyn Rose

The rapid diffusion of artificial intelligence across educational, professional, and civic settings has unsettled long-standing assumptions about how people learn, collaborate, and make sense of complex problems. This plenary symposium invites the society to reimagine this disrupted landscape through the lens of human–AI complementarity: the idea that humans and AI systems can contribute distinct, interdependent strengths that enable new forms of learning and collective capability. Bringing together scholars whose work spans sociotechnical design, collaborative learning, cognitive augmentation, and AI ethics, the session explores how complementary partnerships can repair fractured practices, redistribute expertise, and open pathways for more expansive participation.

The symposium centers on three intertwined questions: (1) How does complementarity reframe disruption (not as loss, but as an opportunity) to rethink core learning processes and relationships? (2) What models of hybrid collaboration reveal when and how AI can amplify, rather than overshadow, human judgment, creativity, and agency? (3) How must learning environments, tools, and policies be reconfigured to cultivate equitable and trustworthy human–AI interdependence? Through empirical cases, design exemplars, and forward-looking provocations, this session offers a vision for reconstituting learning ecologies where disruption becomes a catalyst for more resilient, participatory, and humane futures.

 © ISLS

Long Papers

 © ISLS

 © ISLS

International Society of the Learning Sciences

Designing Multi-Agent AI Systems for Collaborative Learning: Strategic Questioning Patterns and Cognitive Development in Large-Scale Online Learning

Xiaojie Niu, Institute for Advanced Studies in Education, Shandong University, xiaojie.niu@sdu.edu.cn
Jingjing Zhang, Faculty of Education, Beijing Normal University, jingjing.zhang@bnu.edu.cn

Abstract: As large-scale online learning environments increasingly integrate AI technologies to support collaborative learning, understanding how learners strategically interact with different AI agent roles becomes critical for effective system design. Grounded in Bloom's Taxonomy and Graesser and Person's question taxonomy, this study examines questioning strategies employed by 411 doctoral students interacting with three distinct AI agents—Teacher, Assistant, and Companion—in a large-scale online course, and compares cognitive development trajectories between AI-mediated and human peer dialogues. Using clustering analysis on 224 AI-directed questions and sequential analysis of cognitive levels coded via Bloom's Taxonomy, we identified three strategic patterns: Teacher-Focused Quick-Response Strategy (35.7%), Assistant-Focused In-Depth Strategy (27.7%), and Teacher-Focused In-Depth Strategy (36.6%). AI-involved dialogues showed higher cognitive-level variability (average progress of 0.48) compared to human dialogues (0.02), suggesting AI agents accelerate cognitive transitions while human interactions provide stability. These findings illuminate design principles for multi-agent systems that balance AI efficiency with human collaborative depth.

Introduction

The integration of multi-agent AI systems in large-scale online learning environments represents a promising approach to addressing persistent challenges of learner isolation and limited interaction. As platforms serving millions of learners globally face completion rates below 10% (Shah, 2024), AI agents powered by large language models offer role-specific support—teachers for explanatory depth, assistants for procedural guidance, and companions for peer-like encouragement. These multi-agent architectures mirror the diverse social roles present in traditional collaborative learning contexts, potentially enabling scalable support for collaborative knowledge construction in massive open learning environments.

Despite growing adoption of AI agents in large-scale collaborative learning, two critical knowledge gaps limit our understanding of their educational impact. First, while research has examined general AI chatbot usage patterns (Goel & Polepeddi, 2021), we lack empirical evidence on how learners strategically differentiate their questioning approaches across role-specific AI agents. Understanding such strategic differentiation matters because questioning behavior reflects learners' metacognitive awareness of what knowledge they need and from whom they might best obtain it. If learners treat all agents uniformly, multi-agent architectures may offer no advantage over single-agent systems; if learners differentiate strategically, this informs how designers should specialize agent roles to scaffold distinct learning needs. Second, existing studies comparing AI-mediated and human peer interactions focus primarily on learning outcomes (Yan et al., 2024), but fail to characterize the dynamic cognitive progression patterns within dialogues. Understanding trajectory differences is pedagogically important because it can reveal whether AI scaffolding promotes sustained, deep engagement or merely accelerates surface-level transitions—information critical for deciding when to deploy AI support versus human collaboration in instructional design.

Specifically, we operationalize cognitive development as shifts in the cognitive level of learner discourse, coded using Bloom's revised Taxonomy (Krathwohl, 2002), across sequential turns within a dialogue. This operationalization captures the dynamic trajectory of thinking within interactions rather than static pre-post outcomes. We examine two research questions: (1) What role-differentiated questioning strategies do learners employ when interacting with multiple AI agents? (2) How do cognitive level trajectories differ between AI-mediated dialogues and human peer dialogues in collaborative knowledge construction? Using authentic data from 411 doctoral students engaged in collaborative knowledge construction in a large-scale online course, we employed clustering analysis to identify questioning strategies across AI Teacher, Assistant, and Companion roles, and sequential analysis to compare cognitive level sequences coded via Bloom's Taxonomy between AI and human dialogues.

 © ISLS

Related work

Multi-agent AI systems in collaborative learning

Multi-agent AI systems provide role-differentiated support that mimics diverse educational actors, addressing isolation in asynchronous environments by allowing learners to tailor queries to specific agent roles (Burton et al., 2024). Research emphasizes how large language models enhance collective intelligence through role-specialized interactions, though calls persist for granular analysis of query patterns across agents (Tian et al., 2022). Studies have visualized AI applications and identified clusters where multi-agent systems facilitate personalized pathways, yet understanding of how learners strategically select agents for different question types—such as goal-oriented queries to teachers versus instrumental ones to assistants—remains limited. This distinction matters because question type reflects the cognitive function learners seek: goal-oriented questions target conceptual understanding, while instrumental questions seek procedural guidance, and the alignment between question type and agent role can either support or hinder learning efficiency.

Clustering analysis has emerged as a powerful tool in learning analytics for uncovering behavioral patterns in AI-augmented environments. Studies applying lag sequential and epistemic network analysis to collaborative problem-solving with AI have clustered behaviors into cognitive phases, identifying distinct interaction patterns correlating with performance (An & Zhang, 2024). Extensions to multi-agent contexts using sequence mining reveal that short-answer verification queries directed to assistants and long-answer causal exploration queries to teachers predict progression to higher cognitive levels in subsequent discourse (Li et al., 2024). These findings highlight clustering's potential to reveal strategic differentiation in multi-agent querying.

Cognitive trajectories in AI-mediated learning

Cognitive levels in post-query dialogues, evaluated through Bloom's Taxonomy (Krathwohl, 2002), reveal notable differences between AI and human interactions. Bloom's revised Taxonomy classifies cognitive processes into six hierarchical levels—remembering, understanding, applying, analyzing, evaluating, and creating—providing a systematic framework for coding the cognitive demands of discourse. AI often facilitates higher-order processes through structured responses, while humans provide relational depth enhancing creativity. Studies analyzing trajectory paths in AI-tutoring sessions show dialogues escalating to analyzing and evaluating levels via scaffolded questions, contrasting with human-peer exchanges remaining at understanding levels due to variable expertise (Chen et al., 2022). Recent research introduces a cognitive paradox framework integrating Cognitive Load Theory with Bloom's Taxonomy, explaining AI's dual impact: while demonstrating a capacity for cognitive enhancement through adaptive scaffolding, AI simultaneously introduces risks of cognitive offloading that may erode critical thinking (Wang et al., 2024).

Comparative research finds generative AI agents with scaffolding significantly enhance comprehension—aligning with higher Bloom levels—compared to passive or human-only conditions (Yan et al., 2024). Network analysis on group interactions reveals AI's cognitive contributions surpass humans in structured exchanges but lag in spontaneity, highlighting trajectory differences (Feng, 2025). However, concerns about potential negative effects emerge, with research noting higher cognitive performance in AI dialogues but identifying metacognitive laziness risks, underscoring the need for hybrid models balancing efficiency and depth (Abbas et al., 2024). Identifying trajectory differences between AI-mediated and human dialogues is theoretically significant because it illuminates the mechanisms through which each modality supports or constrains cognitive progression, informing decisions about when and how to integrate AI agents within collaborative learning designs.

Methods

Context and participants

This study utilized data from a large-scale general education course titled "China's Educational Reform and Development" implemented in spring 2025 at a research university in China. The course structure included eight thematic modules delivered through 24 videos spanning 16 credit hours, with assessment comprising 60% ongoing performance—including 15% discussion participation—and 40% final examination. The course adopted a large-scale collaborative learning approach where learners engaged in ongoing collaborative knowledge construction activities through the platform's discussion areas around course video content, including asking questions, sharing perspectives, responding to peers, and interacting with AI agents. Participants consisted of 411 first-year doctoral students who provided informed consent for research use of their learning data. The study was approved by the institutional review board.

This study employed an observational design analyzing naturally occurring interaction patterns. All 411 students had equal access to both AI agents and human peers throughout the course. Thus, both AI-involved and human-only dialogues could be generated by the same students, and the comparison reflects differences in interaction modality rather than between-subjects differences. We acknowledge that students who chose to interact with AI may differ from those who did not, representing a limitation inherent to naturalistic observation in authentic educational settings.

Platform and AI system design

The study was conducted using the self-developed Shidao Cloud online learning platform (Version 4.0), featuring video playback, notes and comments areas, and navigation tools. As illustrated in Figure 1, the platform interface consists of three main components: a video playback area providing functions such as video viewing, timeline navigation, and variable-speed playback; a notes and comments area where learners can express viewpoints and questions and engage in replies; and a navigation tools area for personal information settings and chapter directory access.

Figure 1

Shidao Cloud Platform Interface Design



The platform implemented a multi-agent AI system featuring AI Teacher, AI Assistant, and AI Companion roles. As shown in Figure 2, learners could interact with different AI agents by typing "@" followed by selecting the desired agent role, then inputting questions to receive role-specific responses powered by the DeepSeek model. Each agent was equipped with role-specific knowledge bases and structured prompts. AI Teacher provided systematic knowledge and deep analysis using course materials and reading resources. AI Companion offered peer support and encouragement with access to previous learners' reflections. AI Assistant handled course management and platform guidance using comprehensive course documentation.

Data collection and analysis

Data collection encompassed 91,797 comments compiled into 60,826 discussion sequences. Across the full dataset, AI-involved sequences had an average length of 2.89, while human-only sequences averaged 1.82, reflecting the typically brief nature of most online discussion exchanges. To ensure meaningful cognitive trajectory analysis, we filtered for sequences with length ≥ 3, as meaningful cognitive progression requires at least three sequential exchanges to exhibit a trajectory pattern (i.e., a starting state, potential transition, and ending state). This filtering resulted in 7,028 sequences containing 21,084 comments. Of these, AI-involved sequences constituted 2.3% (n=163 sequences containing 853 comments), while human-only sequences constituted 97.7% (n=6,865 sequences containing 59,973 comments). The relatively small proportion of AI-involved interactions

reflects the voluntary nature of AI engagement in the course design. Figure 3 illustrates the research design framework.

Figure 2

Multi-Agent Application Interface

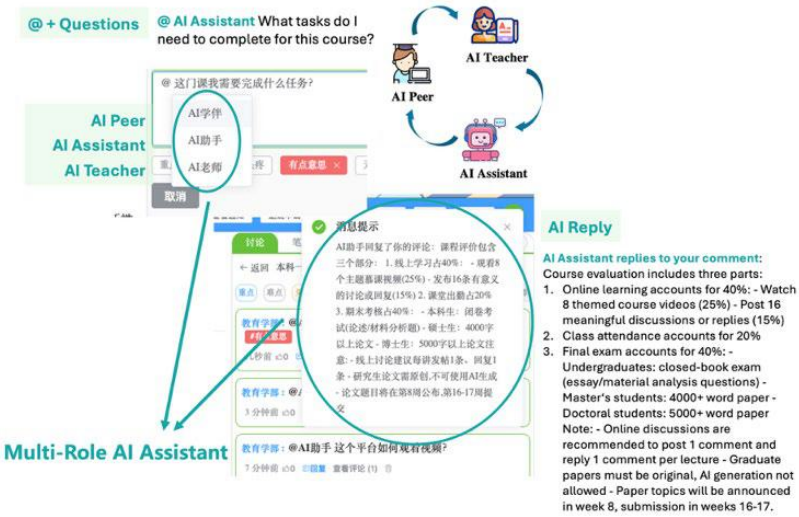
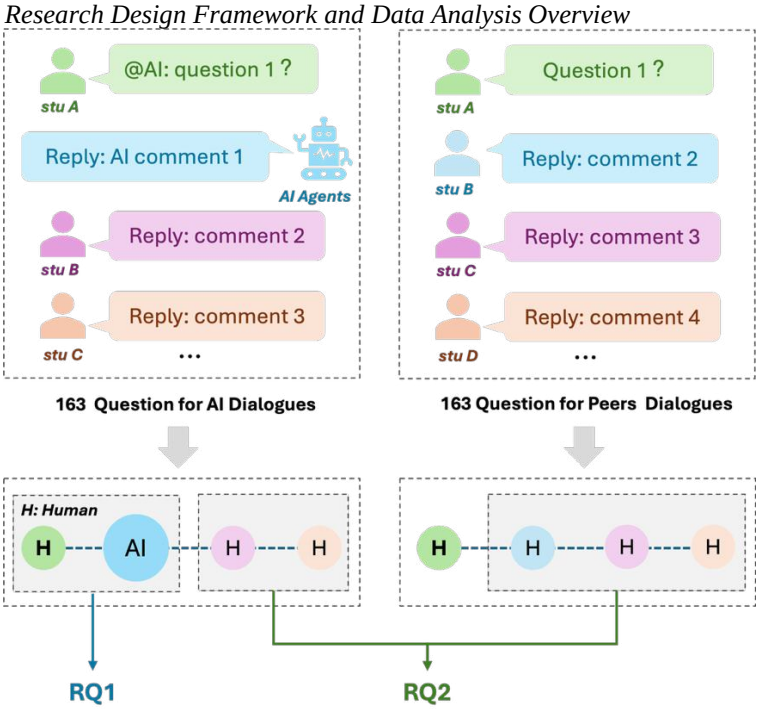


Figure 3

Research Design Framework and Data Analysis Overview



To address the first research question, we conducted K-means clustering analysis on 224 AI-directed questions extracted from the 163 AI-involved sequences. 'Valid' questions were defined as student-initiated queries explicitly directed at a specific AI agent (excluding greetings, follow-up acknowledgments, and off-topic remarks). The unit of analysis for clustering was individual questions, characterized by two features: the target AI agent and question content coded using Graesser and Person's (1994) question taxonomy. This taxonomy categorizes questions into types including concept completion, feature specification, instrumental/procedural, interpretation, and others based on the expected answer format and cognitive demand. We evaluated K values ranging from 2 to 6, computing silhouette scores for each candidate to identify the optimal number of clusters.

The optimal K value was determined through silhouette score optimization. The silhouette score ranges from -1 to 1, where values closer to 1 indicate that data points are well-matched to their own cluster and poorly matched to neighboring clusters; values above 0.5 are generally considered indicative of reasonable clustering structure. Our analysis yielded K=3 with a silhouette score of 0.576, indicating adequate cluster separation.

For the second research question, we conducted sequential analysis of cognitive level sequences in 163 AI-involved dialogues and 163 randomly sampled human-only dialogues. Random sampling was employed for human-only dialogues to create a comparison set matched in sample size to the AI-involved group, enabling parallel analysis while controlling for potential distortions from unequal group sizes. Cognitive levels were coded via Bloom's revised Taxonomy (Krathwohl, 2002), assigning numerical values to six cognitive levels (1=remembering through 6=creating; 0=irrelevant content, defined as utterances not reflecting identifiable cognitive processing of course content, such as social greetings, platform navigation comments, or off-topic remarks). Cognitive progress was quantified as the difference between the last and first cognitive state in each sequence, with sequences classified as escalating (progress > 0), stable (= 0), or declining (< 0). We examined sequence length distributions, cognitive state distributions, and progression patterns to characterize differences between the two interaction modalities.

Coding procedure

Comment coding utilized automated analysis with GPT-4.1 through the Optimized Large Language Model Coding (OLLM-C) framework, implementing systematic calibration based on human expert feedback (Niu & Zhang, 2026). Initial manual coding by two independent researchers on stratified random samples achieved inter-rater reliability via Cohen's Kappa of 0.82 for Bloom's framework and 0.85 for question taxonomy. Following iterative calibration involving baseline coding, discrepancy analysis, and prompt enhancement, final accuracy assessments reached 91% for Bloom's Taxonomy and 92% for question taxonomy against human-coded gold standards, enabling reliable automated coding of the full dataset.

Results

Strategic patterns in multi-agent questioning

The clustering analysis identified three distinct questioning strategies employed by learners when interacting with AI agents. Among 224 valid AI-directed questions (the unit of analysis; note that individual learners could contribute questions to multiple clusters), AI Teacher received the majority (64.7%, n=145), followed by AI Companion (19.6%, n=44) and AI Assistant (15.6%, n=35). Predominant question types were instrumental/procedural (26.8%, n=60) and interpretation questions (13.8%, n=31), indicating preferences for practical, application-oriented inquiries rather than simple factual verification.

Cluster 0: Teacher-Focused Quick-Response Strategy

This cluster (n=80, 35.7%) demonstrated clear role differentiation with strong preference for AI Teacher (78.8%) supplemented by selective AI Companion use (21.2%), with no AI Assistant utilization. Dominant question types were concept completion (35.0%) and feature specification (30.0%), both requiring specific factual responses. This pattern suggests learners viewing AI Teacher as the primary source for rapid conceptual clarification, exemplified by questions like "When did vocational education first appear in our country?" The absence of complex questions indicates a tendency to reserve AI Teacher interactions for immediate knowledge gaps, reflecting an efficiency-maximizing approach to AI engagement.

Cluster 1: Assistant-Focused In-Depth Strategy

This cluster (n=62, 27.7%) demonstrated multi-agent use with balanced distribution: AI Assistant (56.5%) and AI Companion (43.5%) while not engaging AI Teacher. The predominant question type was instrumental/procedural (41.9%), followed by interpretation questions (14.5%), both requiring elaborate process-oriented responses. This reflects learners directing complex "how-to" questions to AI Assistant's specialized procedural guidance, such as implementation strategies and practical applications. The substantial AI Companion use for similar complex questions suggests these learners view peer-level AI as providing alternative perspectives, reflecting a distributed approach to knowledge construction.

Cluster 2: Teacher-Focused In-Depth Strategy

This cluster (n=82, 36.6%) demonstrated exclusive reliance on AI Teacher (100.0%) for comprehensive educational support across all question types. This cluster showed the highest concentration of complex questions, with instrumental/procedural (41.5%) and interpretation questions (26.8%) dominating. Unlike Cluster 1's

distributed approach, Cluster 2 directed all complex inquiries exclusively to AI Teacher, suggesting a preference for centralized, consistent interaction with a single authoritative agent. This contrasts with Cluster 0, which also favored AI Teacher but for simpler factual questions, whereas Cluster 2 used AI Teacher for deeper analytical inquiries. Notably, Cluster 0's inclusion of AI Companion (21.2%) alongside AI Teacher may reflect learners occasionally seeking peer-level validation after obtaining factual answers, whereas Cluster 2's exclusive Teacher reliance suggests that for complex analytical questions, learners preferred a single authoritative source to maintain coherence rather than seeking supplementary perspectives.

Table 1

Questioning Strategies Across AI Agent Clusters

Strategy Cluster	Size (%)	Primary AI (%)	Top Question Type (%)
Teacher-Focused Quick-Response	80 (35.7)	AI Teacher (78.8)	Concept Completion (35.0)
Assistant-Focused In-Depth	62 (27.7)	AI Assistant (56.5)	Instrumental/Procedural (41.9)
Teacher-Focused In-Depth	82 (36.6)	AI Teacher (100.0)	Instrumental/Procedural (41.5)

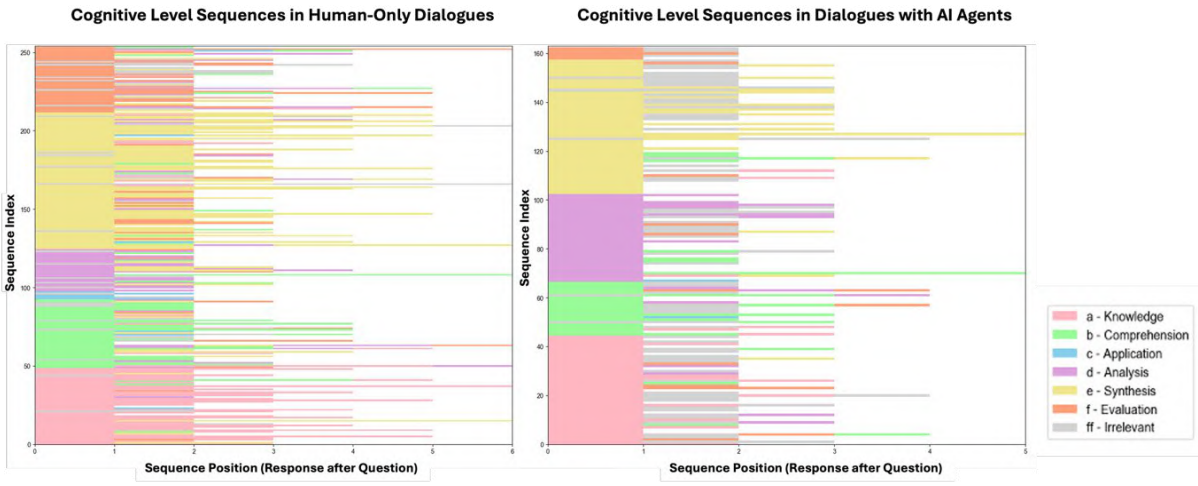
The distribution of questioning strategies demonstrates that learners differentiate their interaction approaches based on agent roles, with distinct question types directed to different agents.

Comparing cognitive level trajectories

The sequential analysis revealed notable differences in cognitive level patterns between AI-mediated and human peer dialogues. Human dialogues demonstrated greater sequence length diversity (mean = 2.66, range 1–6) compared to AI dialogues (mean = 2.07, range 1–5), suggesting that peer-to-peer interactions naturally extend cognitive discourse through sustained reciprocal engagement. Figure 4 provides a comparative visualization of cognitive level sequences between the two interaction modalities.

Figure 4

Comparison of Cognitive Level Sequences: Human-Only vs. AI-Involved Dialogues



As illustrated in Figure 4, human dialogues (left panel) exhibit longer sequences with transitions between cognitive levels characterized by sustained presence at synthesis and analysis levels (represented by purple and yellow bands). In contrast, AI dialogues (right panel) display shorter sequences with more frequent shifts between cognitive levels, including notable amounts of irrelevant content (gray). Quantitatively, the cognitive state distribution in human dialogues showed synthesis comprising 33.0%, knowledge 20.6%, and comprehension 13.5%, indicating sustained higher-order thinking alongside foundational knowledge engagement. AI dialogues showed synthesis at 23.4% but with substantially higher irrelevant content (24.0% vs. 8.2% in human dialogues). This elevated proportion likely reflects the structural characteristics of human-AI interaction. Learners frequently included greetings, agent-selection commands such as "@AI Teacher," and meta-conversational remarks such as expressions of gratitude or acknowledgment that are inherent to the format but carry no cognitive content. Consequently, while AI can stimulate complex thinking, the interaction overhead of engaging AI agents inflates the share of non-cognitive discourse within these sequences.

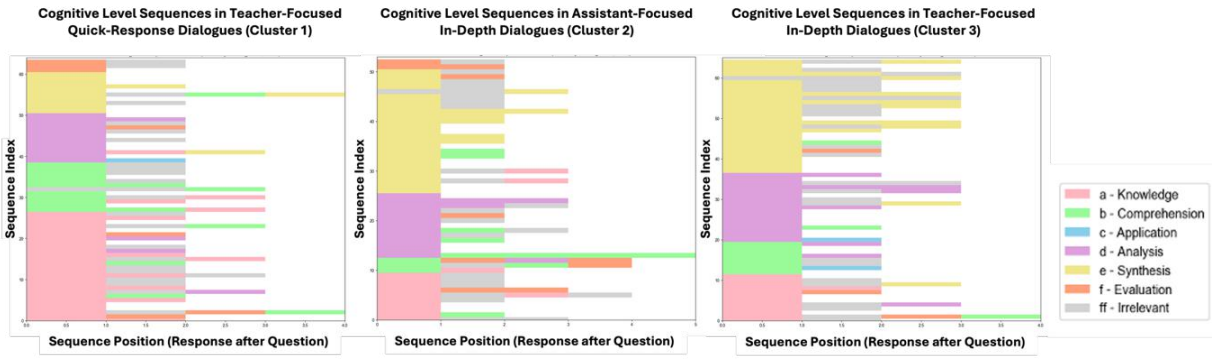
Cognitive progression analysis revealed differences in trajectory patterns. Human dialogues demonstrated near-zero average cognitive progress (0.02) with 60.0% of sequences maintaining stable cognitive

levels, 20.5% showing escalation, and 19.5% declining. This near-zero net change, combined with the high proportion of stable sequences, suggests that human interactions tend to sustain engagement at consistent cognitive levels. AI dialogues showed higher average cognitive progress (0.48) with a different distribution: 35.2% escalating, 46.5% stable, and 18.3% declining. The higher escalation rate in AI dialogues (35.2% vs. 20.5%) suggests that AI interactions may prompt more frequent upward cognitive transitions, potentially through structured scaffolding that guides learners through cognitive hierarchies. However, this should be interpreted cautiously: a higher average progress score does not necessarily indicate deeper learning, as it may also reflect rapid but potentially superficial transitions between levels.

As an exploratory analysis, we examined cognitive trajectory characteristics across the three questioning strategy clusters identified in RQ1 to investigate potential connections between questioning approaches and cognitive outcomes. Figure 5 visualizes these patterns.

Figure 5

Cognitive Level Sequences Across AI Questioning Strategy Clusters



Cluster 0 (Teacher-Focused Quick-Response)

This cluster demonstrated the highest average cognitive progression (0.85) with relatively balanced patterns (42.3% escalating, 42.3% stable, 15.4% declining). The cognitive state distribution emphasized knowledge acquisition (31.9%) and comprehension (16.0%), consistent with the cluster's focus on factual questions. The short sequences concentrated in lower cognitive levels suggest that rapid factual queries to AI Teacher are associated with upward movement from foundational to higher cognitive levels within brief exchanges.

Cluster 1 (Assistant-Focused In-Depth)

This cluster exhibited a complex trajectory pattern with negative average progression (–0.17) and the highest proportion of declining sequences (34.8%). Despite focusing on complex procedural questions, the cognitive sequences showed synthesis activities (28.2%) at initial stages followed by substantial irrelevant content (22.7%) in subsequent positions. One possible interpretation is that learners in this cluster began with analytical thinking but encountered difficulties navigating between different agent capabilities, potentially introducing cognitive overhead that shifted focus away from higher-order processing. However, we note that cognitive decline could also reflect transitions to different but equally valuable cognitive activities not well captured by the Bloom's hierarchy.

Cluster 2 (Teacher-Focused In-Depth)

This cluster showed moderate cognitive progression (0.56) with the highest proportion of stable sequences (52.0%) and lowest decline rate (12.0%). The cognitive state distribution emphasized synthesis (33.1%) and analysis (20.2%), with longer sequences showing sustained higher-level cognitive engagement. The pattern suggests that centralized interaction with a single AI agent may support more coherent cognitive engagement compared to distributed multi-agent approaches, though this interpretation warrants further investigation with controlled designs.

Discussion

This study illuminates how learners navigate multi-agent AI systems and how these interactions shape cognitive level trajectories compared to human peer dialogues. The identification of three questioning strategies extends prior research on role-differentiated AI use (Burton et al., 2024). Unlike earlier broad categorizations (Long & Magerko, 2023), our clustering approach reveals alignment between question types and agent roles: efficiency-

oriented factual queries directed primarily to AI Teacher versus procedural questions distributed across Assistant and Companion. While Clusters 0 and 2 both favored AI Teacher, they did so for qualitatively different purposes—factual clarification versus deep analytical inquiry—suggesting that role differentiation operates at the level of question complexity rather than merely agent selection. This finding complicates a simple narrative of strategic agent differentiation. Rather than primarily distinguishing between agents, learners appear to develop distinct interaction modes with a preferred agent, adapting their questioning depth to match different learning needs within a single agent relationship. For multi-agent system design, this implies that the cognitive versatility of individual agents may matter as much as the diversity of available roles.

The cognitive level analysis quantifies AI's impact using Bloom's Taxonomy. The higher variability in AI dialogues (average progress 0.48) compared to human dialogues (0.02) is consistent with findings from intelligent tutoring systems where structured scaffolding accelerates transitions to higher cognitive levels (Cukurova et al., 2019). Yet the negative progression in the Assistant-Focused In-Depth cluster (–0.17) reveals that multi-agent use may introduce coordination challenges, echoing difficulties noted in distributed AI systems (Chu et al., 2022). Human dialogues' stability (60.0% stable sequences) aligns with meta-analyses highlighting human interactions' strength in fostering sustained engagement through diverse perspectives (Choi et al., 2024). However, we acknowledge that stability could reflect either deep processing or stagnation, and without direct learning outcome measures, we cannot definitively determine which interpretation applies.

Recent evidence from randomized controlled trials demonstrates that well-designed AI tutoring systems can achieve substantial learning gains (Kestin et al., 2025). However, systematic reviews reveal concerning over-reliance patterns (Wang et al., 2024). These align with our observation that while AI dialogues show higher escalation rates, they also exhibit greater irrelevant content and variability. The Teacher-Focused In-Depth strategy's pattern of maintaining cognitive coherence (52.0% stable sequences) suggests centralized AI architectures may be preferable for sustained cognitive engagement, providing consistent scaffolding without the coordination complexity of multi-agent interactions.

From a CSCL(Computer-Supported Collaborative Learning) perspective, these findings raise important questions about the collaborative dimension of AI-mediated learning. While multi-agent systems can provide role-differentiated support at scale, the collaborative knowledge construction between human learners—characterized by sustained engagement and balanced cognitive state distributions—appears to offer unique benefits not easily replicated by AI agents. Future work should more explicitly theorize how individual learner-AI interactions contribute to collective knowledge building within the broader learning community.

Several limitations should be noted. First, the relatively small proportion of AI-involved interactions (1.4%) may limit generalizability; students who voluntarily engaged with AI may differ systematically from those who did not. Second, our observational design cannot establish causal relationships between interaction modality and cognitive outcomes. In particular, the higher cognitive escalation observed in AI dialogues may reflect characteristics of students who chose to engage with AI, such as higher motivation or stronger prior knowledge, rather than effects of the AI interaction itself. Third, without direct learning outcome measures, interpretations regarding cognitive "development" are limited to within-dialogue trajectory patterns. Fourth, the 224 AI-directed questions, while sufficient for exploratory clustering, represent a modest sample that may not capture the full range of questioning behaviors. Future research employing experimental designs with larger AI interaction samples and linked outcome measures would strengthen these findings.

Conclusion

This study provides analysis of learner interactions with multi-agent AI systems in online education, identifying three questioning strategies that reflect differentiated approaches to AI role capabilities. While AI dialogues show higher rates of cognitive level escalation through structured scaffolding, they also introduce variability and irrelevant content compared to human interactions' stability. The exploratory finding that centralized single-agent interaction may support more coherent cognitive engagement than distributed multi-agent approaches offers a counterintuitive but practically valuable insight for system design.

For practice, results offer guidance for multi-agent system design, suggesting that systems should consider providing streamlined interfaces that reduce the cognitive overhead of navigating multiple agents. Educators can implement hybrid models combining AI's scalability with human collaborative elements through structured alternation between AI and peer discussions. Future research should investigate longitudinal effects across disciplines and cultural contexts using experimental designs, and should explicitly examine how AI-mediated individual interactions contribute to collective knowledge construction within CSCL frameworks.

References

Abbas, M., Jam, F. A., & Khan, T. I. (2024). Is it harmful or helpful? Examining the causes and consequences of generative AI usage among university students. *International Journal of Educational Technology in Higher Education, 21*(1), 10.

An, S., & Zhang, S. (2024). Effects of ability grouping on students' collaborative problem solving patterns: Evidence from lag sequence analysis and epistemic network analysis. *Thinking Skills and Creativity, 51*, 101453.

Burton, J. W., Stein, M. K., & Jensen, M. B. (2024). How large language models can reshape collective intelligence. *Nature Human Behaviour, 8*, 1821–1831.

Chen, G., Kumar, V., & Grešković, M. (2022). A systematic review of artificial intelligence techniques for collaborative learning over the past two decades. *Computers & Education: Artificial Intelligence, 3*, 100097.

Choi, S., Lee, S. W., Lee, H., & Lee, D. (2024). Examining AI use in educational contexts: A scoping meta-review and bibliometric analysis. *International Journal of Artificial Intelligence in Education*. Advance online publication.

Chu, S.-T., Hwang, G.-J., & Tu, Y.-F. (2022). Artificial intelligence-based robots in education: A systematic review of selected SSCI publications. *Computers and Education: Artificial Intelligence, 3*, 100091.

Cukurova, M., Kent, C., & Luckin, R. (2019). Systematic review of research on artificial intelligence applications in higher education—where are the educators? *International Journal of Educational Technology in Higher Education, 16*(1), 1–27.

Feng, S. (2025). Group interaction patterns in generative AI-supported collaborative problem solving: Network analysis of the interactions among students and a GAI chatbot. *British Journal of Educational Technology, 56*(2), 2125–2145.

Goel, A., & Polepeddi, L. (2021). *Jill Watson: A virtual teaching assistant for online education*. Georgia Institute of Technology Technical Report.

Graesser, A. C., & Person, N. K. (1994). Question asking during tutoring. *American Educational Research Journal, 31*(1), 104–137.

Kestin, G., Miller, K., Klales, A., Sun, W., Vemu, K., Shultz, G., Crosswell, P., & Mazur, E. (2025). AI tutoring outperforms in-class active learning: An RCT introducing a novel research-based design in an authentic educational setting. *Scientific Reports, 15*(1), 2345.

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into Practice, 41*(4), 212–218.

Li, Y., Xie, H., & Zhong, J. (2024). Investigating the utilization and impact of large language model-based intelligent teaching assistants in flipped classrooms. *Computers in Human Behavior, 152*, 108010.

Long, D., & Magerko, B. (2023). Artificial intelligence (AI) learning tools in K-12 education: A scoping review. *Journal of Computers in Education, 11*(1), 215–252.

Niu, X., & Zhang, J. (2026). Enhancing automated text coding in online learning research: A systematic calibration framework for large language models. *IEEE Transactions on Learning Technologies, 19*, 63–74.

Shah, D. (2024). By the numbers: MOOCs in 2024. https://www.classcentral.com/report/moocs-stats-2024/

Tian, H., He, W., Bai, X., & Chen, Y. (2022). Mapping artificial intelligence literacy in online learning research: A bibliometric analysis. *International Review of Research in Open and Distributed Learning, 23*(4), 1–25.

Wang, H., Zhang, C., Guo, J., Yang, Z., & Sun, Z. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. *Smart Learning Environments, 11*(1), 1–23.

Yan, L., Martinez-Maldonado, R., & Gasevic, D. (2024). The effects of generative AI agents and scaffolding on enhancing students' comprehension of visual learning analytics. *Computers & Education, 232*, 105322.

International Society of the Learning Sciences

Structural Signals and Role Archetypes: The Emergence of Roles in a Distributed Epistemic Game

Jenna Matthews, Utah State University, jenna.matthews@usu.edu

Abstract: Roles in collaborative learning are often scripted or assigned; in distributed epistemic games (DEGs), however, they emerge through interaction over time. This study analyzes the first Polymath Project, a large-scale mathematical collaboration, to show how roles crystallize in practice. Using canonicalized epistemic move sequences, temporal distance metrics, and cross-method clustering, I identify two converging empirical signals. First, participant distance matrices display visible "plaid" structures—evidence that contributors gravitated into role-like repertoires even before clustering. Second, across five analytic pipelines, three participant groupings ('consistent companionships') persisted: Proof Builders, Paired Refiners, and Process Integrators. Their recurrence across methods supports viewing them as durable role archetypes. Conceptually, the findings show that roles in DEGs are emergent and self-organizing; methodologically, they demonstrate how structural and reflexive analysis (via AIRR) can jointly surface roles-in-practice, advancing methods for studying collective intelligence and informing the design of systems that support dynamic role ecologies.

Introduction

In large, open collaborations, participants' patterned contributions—conjecturing, testing, refining, coordinating—aggregate into collective trajectories of knowledge-building. Unlike scripted classroom settings, DEGs such as Polymath have no assigned roles; instead, recurrent repertoires of action become visible over time (Kollar et al., 2006; Strijbos & De Laat, 2010). I investigate how such repertoires emerge and stabilize, making two contributions:

1. Empirical: I provide converging evidence that role structures are detectable directly in temporal sequences. Even prior to clustering, pairwise distance heatmaps reveal plaid patterns—alternating light/dark blocks indicating pockets of strong repertoire similarity interleaved with sharp divergence. Across five clustering pipelines, three consistent companionships (Proof Builders, Paired Refiners, Process Integrators) persist, warranting the status of archetypes.
2. Methodological: I articulate a transparent sequence-analytic workflow and introduce a reflexive reliability framework (AIRR) that treats human–AI analytic divergence as data, enabling traceable refinement of interpretive decisions through (a) canonicalization of epistemic moves, (b) multiple distance measures (DTW/edit), and (c) cross-method validation (Berndt & Clifford, 1994; Levenshtein, 1966; Müller, 2011). The workflow yields artifacts (distance matrices, cluster rationales, refinement notes) that make interpretive evolution inspectable (Chen, Knight, & Wise, 2018).

DEGs sustain layered role ecologies—proof engines, evaluative dyads, and process anchors—each playing complementary parts in collective progress (Arazy et al., 2016; Suthers, 2006; Wenger, 1998).

Related work

In most CSCL studies, roles are carefully scripted to support productive collaboration—for example, by assigning learners to act as summarizers, explainers, or critics (Kollar et al., 2006; Strijbos & De Laat, 2010; Strijbos & Weinberger, 2010; Saqr & López-Pernas, 2022). Such scaffolds make participation visible and equitable, but they do not capture how roles evolve once interaction becomes open-ended. By contrast, large-scale collectives such as Wikipedia or citizen science projects show that recognizable roles can emerge spontaneously from repeated contribution patterns (Arazy, Ortega, Nov, Yeo, & Balila, 2016; Masters et al., 2016; Watson & Floridi, 2016). Participants differentiate themselves not through assignment but through how they work, how often they engage, and the epistemic moves they repeat or combine (De Wever, Keer, Schellens, & Valcke, 2010; Dowell, Lin, Godfrey, & Brooks, 2020; Hecker & Taddicken, 2022). Distributed epistemic games extend this dynamic: contributors take on characteristic repertoires of action that can be traced across time rather than predefined categories.

Recent research in learning analytics has emphasized the temporal dimension of participation—how sequences of discourse moves relate to coordination and learning (Chen et al., 2018; Chiu, 2018; Dowell et al., 2020; Rosé et al., 2008). Building on this line of work, I analyze ordered move sequences from Polymath participants to uncover evolving repertoires of epistemic practice. I employ both edit distance, which captures

 © ISLS

local substitution and insertion patterns (Levenshtein, 1966), and Dynamic Time Warping (DTW), which aligns sequences even when their pacing differs (Berndt & Clifford, 1994). Together, these measures show similarity in rhythm and structure across participants, exposing the latent architecture of collective reasoning.

These perspectives intersect with studies of role ecologies and collective intelligence, where large-scale collaborations are described as layered networks of complementary functions—coordinators, challengers, integrators, synthesizers (Arazy et al., 2016; Watson & Floridi, 2016; Wenger, 1998). My contribution is to demonstrate that such interdependent roles can be detected empirically: their traces appear as stable structures within distance matrices and recur across independent clustering methods. This mirrors insights on distributed expertise in knowledge-building communities (Scardamalia & Bereiter, 1994; Sawyer, 2005).

This framing positions the Polymath Project as an ideal case for empirical analysis. Because every contribution was timestamped and textually available, it offers a uniquely transparent record of how epistemic moves accumulate into patterned repertoires (Gowers, 2009; Cranshaw & Kittur, 2011). To trace those patterns, I adopted a mixed computational and interpretive strategy: refining data collaboratively through AI Reliability through Refinement (AIRR) and analyzing participants' move sequences using both temporal and structural perspectives (Matthews et al., 2024). The following section outlines how the dataset was prepared, canonicalized, and analyzed to identify the structural signals and consistent companionships that define emergent role archetypes.

Methods

My methodological goal was to make the emergence of roles empirically visible without reducing them to pre-labeled categories. To do so, I combined computational sequence analysis with a reflexive reliability framework (AIRR), in which human–AI interpretation is treated as a structured, traceable analytic process. The analysis proceeded in two complementary strands. First, a quantitative pipeline identified recurring patterns of epistemic activity through canonicalized move sequences, distance metrics, and clustering. Second, a qualitative layer, conducted through the AIRR workflow, documented how analytic decisions evolved, and how disagreements between human and AI readings were resolved. This dual approach provided both the reproducibility of computation and the transparency of interpretive traceability (Chen et al., 2018; Holstein, Aleven, & Rummel, 2020; Woolf, Lane, Chaudhri, & Kolodner, 2013).

The data

The analysis draws on the Polymath Project (2009), an open, blog-based mathematical collaboration initiated as an experiment in "massively collaborative mathematics" (Gowers, 2009; Cranshaw & Kittur, 2011). Designed for public contribution in threaded comments, the project left a rich digital record spanning exploration to solution. For this study, I constructed a canonicalized dataset comprising approximately 118 distinct participants and their posts, each annotated with author identity, timestamp, thread context, and coded epistemic moves:

- TAP — Think-Aloud Problem-solving or working in progress
- Evaluate — Testing, checking, or verification
- Idea — Proposal, conjecture, or heuristic introduction
- Meta — Coordination and process management
- Clarify — Communication or clarification moves

This representation treats discourse as sequences of epistemic actions rather than isolated messages, allowing temporal and structural analysis of how reasoning evolved.

Ethical considerations

All data analyzed are publicly available. To preserve privacy and scholarly integrity, findings are reported at the level of role archetypes and aggregated participation patterns. No new identifying information is introduced beyond what is public and essential for reproducibility.

Data preparation

Preparing the dataset required reconciling the openness of Polymath's participation with the precision needed for temporal and structural analysis. Two refinements were essential: consolidating participant identities and canonicalizing epistemic moves. Because contributors used a mixture of personal names and online handles, I merged clear duplicates (e.g., "Tao" and "Terence Tao") while retaining uncertain cases as distinct to avoid false conflation. Each post was coded by its primary epistemic function using the five canonical move types above. Multi-word codes (e.g., Meta/Coordination) were remapped into single tokens (Meta) to prevent artificial inflation of n-grams during sequence extraction. These choices align with calls for transparent analytic decisions in

temporal analytics (Chen et al., 2018) and with reflexive, hybrid adaptivity in human–AI workflows (Holstein et al., 2020).

The preparation process exemplified AIRR: early discrepancies between human and AI tokenization revealed hidden boundary rules (e.g., compound labels preserved as single epistemic units). Rather than treating mismatches as errors, they were documented and debated, refining the canonical map through iterative human–AI negotiation (Matthews et al., 2024). With identities consolidated and moves canonicalized, the dataset was ready for sequence-level analysis.

Analytic pipeline

Each participant's contributions were represented as an ordered sequence of canonicalized move codes—TAP, Evaluate, Idea, Meta, Clarify—preserving temporal flow. To quantify similarity between trajectories, I employed two complementary distance measures: Levenshtein edit distance for local substitution/insert/delete sensitivity (Levenshtein, 1966) and Dynamic Time Warping for tempo-invariant alignment (Berndt & Clifford, 1994). Pairwise distance matrices were visualized as heatmaps. The resulting plaid structures—interwoven bands of similarity—offered a first visual signal that participants gravitated into role-like repertoires without predefinition. To test robustness, I conducted five independent clustering analyses with varying parameters, inclusion thresholds, and linkage methods (Müllner, 2011), then compared stability across runs to identify consistent companionships—groupings that persisted across all analytic variations.

At each stage, computational outputs were reviewed through the AIRR lens, pairing algorithmic sensitivity with human interpretive judgment (Chen et al., 2018; Holstein et al., 2020). Stability across methods served as the strongest evidence of role robustness.

Reliability and reflexivity

Traditional inter-rater reliability assumes that agreement among coders signals validity; that model underrepresents the interpretive complexity of evolving, hybrid analyses. AIRR reframes reliability as iterative, documented negotiation between human and AI interpretations (Matthews et al., 2024). Divergence is treated as data: disagreements (e.g., whether to collapse repeated moves) triggered refinement cycles logged with reasoning and outcomes. Artifacts—comparison tables, refinement logs, decision memos—form an audit trail, aligning with calls for interpretive accountability in temporal analytics (Chen et al., 2018) and hybrid adaptivity (Holstein et al., 2020; Woolf et al., 2013).

All key AIRR stages generated material artifacts: comparison tables, refinement logs, and decision memos documenting both human and AI rationales. These artifacts serve as an audit trail, allowing readers to trace how specific analytical judgments were made and how refinement shaped the final dataset. In this sense, AIRR transforms reliability from a static metric into a reflexive workflow that captures the epistemic life of the analysis itself.

This approach also leverages the distinct strengths of human and AI reasoning. The AI offers scalability and pattern sensitivity across vast textual data, while the human analyst provides theoretical grounding and contextual understanding. Reliability thus emerges not from replication but from reconciliation—the point at which machine regularity and human interpretation converge through iterative dialogue.

Through AIRR, reliability in this study became evidence of process transparency rather than coder uniformity. It provides a traceable record of how analytic categories were refined, how disagreements were resolved, and how collective understanding was incrementally stabilized—a reflexive stance consistent with the study's broader interest in how coherence itself emerges within distributed systems.

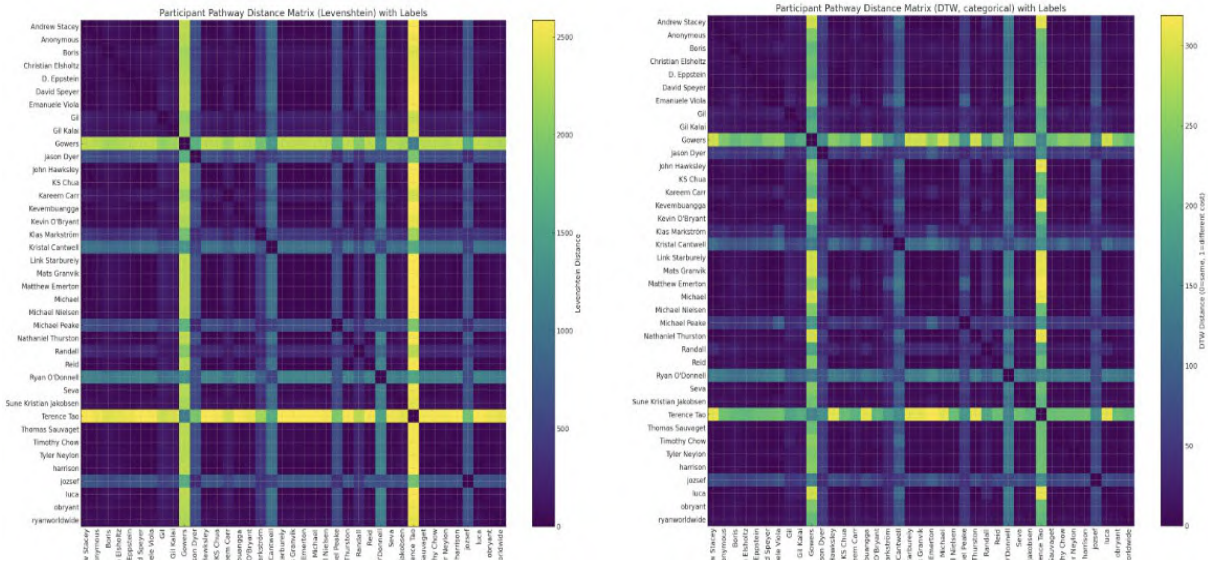
Taken together, these methodological choices established a framework for observing role emergence as both a computational and epistemic phenomenon. The combination of sequence modeling, distance-based similarity measures, and cross-method clustering made it possible to detect durable structural signals within the Polymath collaboration. The AIRR framework added a reflexive dimension, ensuring that analytic reliability emerges through transparent negotiation rather than static agreement. What follows turns from process to pattern—showing how these methods revealed the plaid structures that mark participants' converging repertoires and the consistent companionships that define enduring role archetypes within the distributed epistemic game.

Results

The results of this analysis reveal two complementary forms of evidence for emergent roles in distributed epistemic games. The first is structural, visible in the patterned geometry of participant distance matrices—visual "plaids" that capture how contributors gravitated toward distinct repertoires of epistemic action. The second is robustness-based, demonstrated through the recurrence of specific participant groupings across five independent clustering methods. Together, these patterns provide converging evidence that roles in the Polymath Project

emerged organically from sequences of interaction, leaving measurable traces of coherence in both structure and stability.

Figure 1
Plaid Patterns in Distance Matrices



Structural signals: Plaid patterns in distance matrices

The earliest signal of emergent roles appeared not in the clustering results but within the raw distance matrices themselves as shown in figure 1 above. When participant trajectories were compared pairwise using both Levenshtein and Dynamic Time Warping (DTW) distances, the resulting visualizations revealed a striking plaid structure—a cross-hatched pattern of light and dark bands indicating recurrent pockets of similarity among contributors.

The lighter diagonal blocks correspond to groups of participants whose move sequences followed similar epistemic repertoires, while the darker cross-bands mark sharp divergences between clusters. Rather than a smooth gradient or random dispersion, the matrix displayed alternating ribbons of coherence and contrast, suggesting that contributors naturally coalesced into semi-stable groups with shared tactical rhythms. This emergent structure was consistent across both distance measures, indicating that it was not an artifact of a particular alignment method but a robust signal of underlying organization.

The two distance metrics offered complementary views of these dynamics. Levenshtein distance, sensitive to local substitutions and deletions, emphasized micro-level distinctiveness in participants' reasoning styles—where small shifts in epistemic sequence created sharp contrasts between groups. In contrast, DTW allowed flexible temporal alignment, revealing broader rhythmic affinities even among contributors who participated at different tempos or durations. The persistence of the plaid structure across both measures confirmed that role differentiation was a structural property of the collaboration, not a byproduct of timing or scale.

The plaid pattern represents a structural signal of roles-in-practice—the first indication that roles in the Polymath collaboration crystallized from temporal interaction sequences rather than from external assignment. It captures the collective's latent organization: how participants gravitated into recognizable repertoires of epistemic action that stabilized the discourse. In this sense, the plaid structure serves as a visual metaphor for distributed cognition—revealing how coherence emerges through patterned diversity.

Consistent companionships: Stable archetypes across clustering methods

While the plaid structures revealed role differentiation as a structural pattern, the clustering analyses provided stability evidence—showing that certain participant groupings reappeared across multiple analytic approaches. Across five independent clustering runs—including proportional (Role Arc), temporal (DTW), integrated (Levenshtein–DTW fusion), replication (Jupyter reconstruction), and significance-filtered (Python subset) analyses—three participant clusters remained intact under every variation.

These groupings, which I call consistent companionships, represent enduring role archetypes within the Polymath collaboration. Remarkably, their stability exceeded that of the project's founding hosts: while Gowers

 © ISLS

and Tao anchored several clusters, their membership boundaries shifted across methods, whereas the companionships below persisted unchanged.

Proof builders — This group (Jason Dyer, jozsef, and Michael Peake) formed the collaboration's proof engine. Their sequences were dominated by TAP → Evaluate cycles, reflecting sustained work on formal proofs punctuated by verification. Quantitatively, they produced over 100 posts per member—far above the project average—and sustained participation over 500+ days. Textually, their posts were rich in formal notation and structural argumentation. In the network, they occupied central, bridging positions across multiple threads, serving as connective tissue between problem subspaces.

Paired refiners — Comprising Kristal Cantwell and Ryan O'Donnell, this dyad enacted a refinement role built on iterative evaluation and idea injection. Their contributions displayed dense Evaluate → Evaluate loops punctuated by Idea → TAP transitions, suggesting a reciprocal rhythm of testing and extending conjectures. They were the most central participants in the bipartite participant–thread network (degree ≈ .48, betweenness ≈ .06), reflecting their influence in stabilizing ongoing reasoning and anchoring key problem threads.

Process integrators — Christian Elsholtz and Kevin O'Bryant exemplified process anchoring and oversight. Their lower posting volume (≈15 posts each) contrasted with long participation spans (over a year) and distinctive Eval–Meta–TAP cycles, embedding evaluation within coordination and meta-level reflection. Their textual contributions often linked Polymath discussions to external literature (e.g., Behrend's construction), situating local reasoning within established mathematical frameworks.

Together, these companionships highlight a layered role ecology within distributed collaboration. Proof Builders maintained progress through dense proof cycles, Paired Refiners sustained epistemic stability through evaluative iteration, and Process Integrators maintained coherence through reflective oversight. Each archetype exhibited persistence, internal convergence, and structural differentiation—evidence that emergent roles can be empirically recovered from the temporal traces of epistemic activity.

The convergence of these companionships across independent clustering runs strengthens the claim that emergent roles in distributed epistemic games are both detectable and durable. They are not analytic artifacts but recurring social–cognitive formations—patterns of contribution that stabilize collective inquiry. In this sense, the Polymath Project demonstrates how role archetypes can emerge naturally within distributed epistemic systems, providing structural scaffolds for sustained knowledge-building.

Discussion

The findings from the Polymath Project illustrate that roles in distributed epistemic games emerge not from explicit coordination or assignment but from the patterned sequencing of epistemic activity over time. The convergence of structural signals (plaid patterns) and consistent companionships (stable archetypes) demonstrates that collective inquiry leaves measurable traces of organization even in open, unscripted environments. These results advance theory on role emergence, provide methodological innovations for studying large-scale collaboration, and highlight the potential of reflexive human–AI analysis to make interpretive reasoning itself empirically visible.

Structural signals: Plaid patterns in distance matrices

Findings show that, in open collectives, roles are trajectories of practice—extended sequences of epistemic moves that stabilize through repetition and refinement—rather than static labels (Chen et al., 2018; Chiu, 2018; Dowell et al., 2020). Plaid patterns provide an empirical visualization of these trajectories, capturing latent organization. Coherence emerges from patterned diversity, not uniform participation (Scardamalia & Bereiter, 1994; Sawyer, 2005; Oshima, 2019).

Robustness and ecology of roles

The three archetypes demonstrate structural robustness and layered interdependence (Arazy et al., 2016; Masters et al., 2016; Watson & Floridi, 2016). Proof builders, paired refiners, and process integrators collectively balance innovation and stability, echoing broader accounts of collective intelligence (He et al., 2019; Baltzersen, 2022).

Together, these archetypes form a dynamic ecology that balances innovation and stability—mirroring how collective intelligence systems maintain adaptability without fragmentation. Such role ecologies illustrate the concept of epistemic elasticity across human-only collaborations: the capacity to stretch toward novelty while returning to coherence.

Methodological contributions: Reflexive reliability through AIRR

AIRR reconceptualizes reliability as a traceable refinement process in hybrid analytics, aligning with calls for transparency and human–AI accountability in the design of human–AI systems (Amershi et al., 2019; Chen et al.,

2018; Holstein et al., 2020; Woolf et al., 2013). In the pipeline, potential inconsistencies surfaced by computation became resources for conceptual clarification through documented negotiation (Matthews et al., 2024). AIRR treats human–AI interaction not as a supporting tool but as a designed analytic process whose structure shapes interpretive outcomes (Amershi et al., 2019).

Operationally, AIRR allowed computational methods to surface possible inconsistencies (e.g., tokenization artifacts, sequence distortions), which were then debated and resolved through documented refinement. This reflexive loop transformed potential analytic ambiguity into a resource for conceptual clarification. In doing so, AIRR aligns with emerging calls in CSCL for interpretive accountability in hybrid analytic workflows—making the reasoning process itself a traceable object of study.

Methodologically, AIRR bridges quantitative and qualitative traditions. It pairs the scalability of computation (n-grams, clustering, distance metrics) with the interpretive discipline of qualitative analysis (reflexive memoing, decision logging). The result is an analytic framework that is both reproducible and narratively transparent—a model for future work at the intersection of data-intensive and theory-driven research.

Implications and future directions

The results open several avenues for further research. First, they invite replication across other distributed epistemic games and large-scale collaborations—scientific, educational, or creative—to test whether similar archetypes and structural signals appear. If such plaid structures and consistent companionships recur across contexts, they may represent a generalizable grammar of distributed reasoning.

Second, the findings suggest design implications for systems that support collaborative knowledge-building. Tools could be developed to visualize emergent role patterns in real time, making structural feedback accessible to participants and facilitators alike. Such affordances might encourage reflective awareness of collective dynamics without imposing top-down scripts.

Finally, AIRR's reflexive approach to reliability can extend beyond this study. Future research might formalize refinement cycles, test their generalizability across analytic teams, and integrate them into reproducibility standards for hybrid (human–AI) research.

Synthesis

Coherence in DEGs is not imposed but emerges through recurrent patterns of epistemic action. Plaid structures and consistent companionships show that open, asynchronous collectives develop stable, interlocking repertoires sustaining progress over time. AIRR makes that emergence traceable—not only in the data but in the interpretive process itself.

In this sense, the Polymath Project demonstrates that large-scale collaboration is both a social and epistemic achievement: coherence without script, reliability through refinement.

Conclusion

This study demonstrates that roles in distributed epistemic games emerge through patterned interplay of epistemic moves rather than external assignment. Temporal sequence analysis plus reflexive human–AI refinement revealed two complementary signatures in Polymath: plaid structures visualizing latent coherence and consistent companionships defining durable role archetypes. Large-scale, open collaborations can self-organize into stable ecologies of contribution—sustaining collective intelligence through diversity, not uniformity. Methodologically, AIRR advances transparent, iterative validation in hybrid workflows. Viewed together, these strands point to epistemic elasticity: the capacity of a system to stretch toward novelty, absorb interpretive difference, and return to coherence (He et al., 2019; Baltzersen, 2022). Understanding and supporting that elasticity may be key to designing the next generation of collaborative knowledge systems.

References

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P., Inkpen, K., Teevan, J., Kikin-Gil, R., & Horvitz, E. (2019). Guidelines for human-AI interaction. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.* https://doi.org/10.1145/3290605.3300233

Arazy, O., Ortega, F., Nov, O., Yeo, L., & Balila, A. (2016). Wikipedia's editing dynamics: Toward a quantitative model of participation. *User Modeling and User-Adapted Interaction, 26*(5), 351–386. https://doi.org/10.1007/s11257-016-9171-2

Baltzersen, R. K. (2022). *Cultural-historical perspectives on collective intelligence*. Cambridge University Press.

Berndt, D. J., & Clifford, J. (1994). Using dynamic time warping to find patterns in time series. *KDD Workshop*, 359–370.

Chen, B., Knight, S., & Wise, A. F. (2018). Critical issues in designing and implementing temporal analytics. *Journal of Learning Analytics, 5*(1), 9–28. https://doi.org/10.18608/jla.2018.51.2

Chiu, M. M. (2018). Statistically modelling effects of dynamic processes on outcomes. *Journal of Learning Analytics, 5*(1), 58–67. https://doi.org/10.18608/jla.2018.51.5

Cranshaw, J., & Kittur, A. (2011). The Polymath Project: Lessons from a successful online collaboration in mathematics. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1865–1874). Association for Computing Machinery. https://doi.org/10.1145/1978942.1979213

De Wever, B., Van Keer, H., Schellens, T., & Valcke, M. (2010). Roles as a structuring tool in online discussion groups: The differential impact of different roles on social knowledge construction. *Computers in Human Behavior, 26*(4), 516–523. https://doi.org/10.1016/j.chb.2009.08.008

Dowell, N., Lin, Y., Godfrey, A., & Brooks, C. (2020). Exploring the relationship between emergent sociocognitive roles, collaborative problem-solving skills, and outcomes: A group communication analysis. *Journal of Learning Analytics, 7*(1), 54–71. https://doi.org/10.18608/jla.2020.71.4

Gowers, T. (2009, January 27). Is massively collaborative mathematics possible? *Gowers's Weblog*. https://gowers.wordpress.com/2009/01/27/is-massively-collaborative-mathematics-possible/

He, F., Pan, Y., Lin, Q., Miao, X., & Chen, Z. (2019). Collective intelligence: A taxonomy and survey. *IEEE Access, 7*, 170213–170225. https://doi.org/10.1109/ACCESS.2019.2955281

Hecker, S., & Taddicken, M. (2022). Deconstructing citizen science: Roles. *Journal of Science Communication, 21*(01), A07. https://doi.org/10.22323/2.21010207

Holstein, K., Aleven, V., & Rummel, N. (2020). A conceptual framework for human–AI hybrid adaptivity in education. In *Proceedings of the 21st International Conference on Artificial Intelligence in Education (AIED 2020)* (pp. 240–254). Springer. https://doi.org/10.1007/978-3-030-52237-7_20

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics–Doklady, 10*, 707–710.

Lim, L., Dawson, S., Joksimovic, S., & Gašević, D. (2019). Exploring students' sensemaking of learning analytics dashboards: Where do we go from here? In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge* (pp. 250–259). Association for Computing Machinery. https://doi.org/10.1145/3303772.3303793

Masters, K., Oh, E. Y., Cox, J., Simmons, B., Lintott, C., Graham, G., Greenhill, A., & Holmes, K. (2016). Science learning via participation in online citizen science. *arXiv*. https://doi.org/10.48550/arXiv.1601.05973

Matthews, J., Nguyen, H., & Swanson, H. (2024). Shall we play a game? Distributed games with a generative AI player. *In Proceedings of the 15th International Conference on Computer-Supported Collaborative Learning (CSCL 2024)* (pp. 139–146). International Society of the Learning Sciences. https://doi.org/10.22318/cscl2024.638463

Müllner, D. (2011). Modern hierarchical agglomerative clustering algorithms. *arXiv*. https://arxiv.org/abs/1109.2378

Oshima, J. (2019). Collective knowledge advancement and shared epistemic agency. *Journal of the Learning Sciences, 28*(4–5), 513–550. https://doi.org/10.1080/10508406.2019.1620083

Rosé, C. P., Wang, Y.-C., Cui, Y., Arguello, J., Stegmann, K., Weinberger, A., & Fischer, F. (2008). Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in CSCL. *International Journal of Computer-Supported Collaborative Learning, 3*(3), 237–271. https://doi.org/10.1007/s11412-007-9034-0

Sawyer, R. K. (2005). *The Cambridge handbook of the learning sciences*. Cambridge University Press.

Saqr, M., & López-Pernas, S. (2022). How CSCL roles emerge, persist, transition, and evolve over time: A four-year longitudinal study. *Computers & Education, 189*, 104581. https://doi.org/10.1016/j.compedu.2022.104581

Scardamalia, M., & Bereiter, C. (1994). Computer support for knowledge-building communities. *The Journal of the Learning Sciences, 3*(3), 265–283. https://doi.org/10.1207/s15327809jls0303_3

Strijbos, J.-W., & De Laat, M. F. (2010). Developing the role concept for computer-supported collaborative learning: An exploratory synthesis. *Computers in Human Behavior, 26*(4), 495–505. https://doi.org/10.1016/j.chb.2009.08.006

Strijbos, J.-W., & Weinberger, A. (2010). Emerging and scripted roles in computer-supported collaborative learning. *Computers in Human Behavior, 26(*4), 491–494. https://doi.org/10.1016/j.chb.2009.08.006

Suthers, D. D. (2006). Technology affordances for intersubjective meaning making: A research agenda for CSCL. *International Journal of Computer-Supported Collaborative Learning, 1*(3), 315–337. https://doi.org/10.1007/s11412-006-9660-y

Watson, D., & Floridi, L. (2016). Crowdsourced science: Sociotechnical epistemology in the e-research paradigm. *arXiv*. https://arxiv.org/abs/1610.09485

Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity*. Cambridge University Press.

Woolf, B. P., Lane, H. C., Chaudhri, V. K., & Kolodner, J. (2013). AI grand challenges for education and learning. *AI Magazine, 34*(4), 66–84. https://doi.org/10.1609/aimag.v34i4.2483

International Society of the Learning Sciences

Biased Chatbot Interaction as Preparation for Collaborative Learning

Tanmay Sinha, National Institute of Education, Nanyang Technological University, tanmay.sinha@nie.edu.sg
Kumaran Rajaram, Nanyang Business School, Nanyang Technological University, rkumaran@ntu.edu.sg

Abstract: Twenty-five teachers engaged in individual preparation with chatbots deliberately designed to exhibit biased stances on the learning topic of *AI adoption in higher education* before engaging in collaboration with their human peers. Through intermediate reflections, we probed into how everyone consolidated their learning from the initial *human-AI collaboration*, whether and how this affected reasoning during groupwork, and how everyone consolidated their learning post *human-human collaboration*. Results showed that while teachers recognized and critiqued biases of the chatbots, their group discussions deepened understanding of ethical concerns and the complexity of biased views, leading to a more nuanced and criteria-driven endorsement of AI for education. Overall, our design fostered reflection and perspective-taking, although some teachers remained resistant to changing their views based on feasibility concerns.

Introduction

What can we do to seed productive discourse in learning (Vogel et al., 2017)? How may we increase the likelihood that collaborative learning triggers rich processes like argumentation and mutual regulation (Dillenbourg, 1999) required for *collaborative facilitation* (Nokes-Malach et al., 2015; Rajaram, 2021)? Learning sciences scholarship has long grappled with these central questions as means to design more effective scaffolding for collaboration. We propose that interaction with biased chatbots prior to groupwork can serve as one such scaffold to unlock unique learning affordances. By allowing for generation of partially-formed suboptimal ideas, forcing deep elaboration of the bias, and having learners experience conflict with personal viewpoints, this learning design can create a fertile ground for surfacing internalized tensions and reconciling different, often contradictory viewpoints later. That is precisely the foci of our present work, where by aligning *human-AI* and *human-human* collaboration with principles of *preparation*, *struggle*, and *argumentation*, we create opportunities for teachers to deepen their understanding of the light and dark sides of *AI adoption in higher education*, a critical topic in current discourse.

Theoretical background

Our study is informed by three theoretical lenses from the learning sciences – *individual work as preparation for collaboration* (Lam & Kapur, 2018), *guided failure* (Sinha et al., 2021), and *learning through argumentation* (Asterhan & Schwarz, 2016) – while addressing design gaps in the use of chatbots for collaborative learning. Former work has shown that engaging learners *individually before groupwork* surfaces diverse, partially formed ideas that can enrich collaboration (Lam & Kapur, 2018; Mende et al., 2021; Chen et al., 2024). When this individual work involves structured struggle, such as encountering incorrect or biased information, learners benefit from *guided failure* – grappling with suboptimal ideas and viewpoints along with their underlying epistemic tensions under supportive scaffolding has been shown to foster deeper understanding (Sinha et al., 2021; Sinha & Kapur, 2021; Thorgeirsson et al., 2022). Failure in this sense is not an endpoint but a catalyst, particularly when learners are prompted to question, critique, and refine their reasoning, rather than passively accepting ideas. These preparatory processes set the stage for *argumentation*, in which group members advance and interrogate the credibility of claims, exchange reasons, and consider counterarguments, thereby strengthening both conceptual knowledge through sensemaking and epistemic cognition through an examination of the reliability and credibility of knowledge sources (Chinn & Clark, 2013; Asterhan & Schwarz, 2016).

Meanwhile, recent research on educational chatbots within collaboration has largely positioned them as scaffolding agents that can facilitate collective efficacy (Weng et al., 2024), socio-emotional support, real-time feedback (e.g., interdependence, belongingness), and dialogic engagement (Zhang et al., 2024; Kovari, 2025; Kim & Lim, 2025). The majority of these chatbots are typically integrated into the group discussion itself, functioning as a tutor, mediator or peer, rather than creating a prior safe space for learners to surface their individual reasoning or biases ahead of discussion with human partners. While valuable, this design pattern overlooks the possibility of using chatbots as means of preparation for collaboration, rather than as co-participants in it. Here, our teacher participants individually interacted with biased chatbots before group discussion, which allowed them to surface assumptions (*preparation*), confront partiality (*guided failure*), and subsequently engage in peer *argumentation* informed by prior critical reflection. By shifting the role of chatbots from *"for collaboration"* to *"preparation for*

 © ISLS

collaboration", we examine this under-explored design possibility of using chatbot-mediated individual reflection (*human-AI collaboration*) to prime richer, more critical peer argumentation (*human-human collaboration*).

Method

Participants

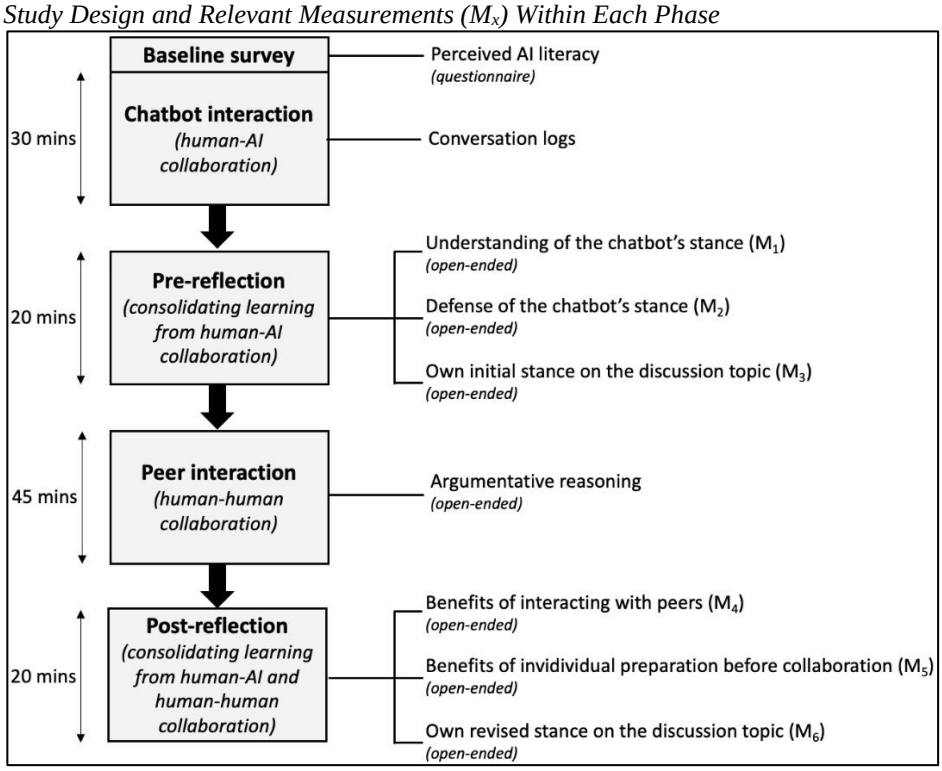
We engaged N = 25 primary, secondary, and postsecondary school teachers from Singapore as part of regular coursework at a teacher training institute. The majority of participants had completed a postgraduate diploma in education. Informed, written consent was obtained from n = 20 of the 25 participants. The study was approved by the ethics commission of Nanyang Technological University (IRB-2024-1081).

Study design

We introduced a novel multi-phase study design (Figure 1), where everyone started out by self-reporting their perceived baseline levels of AI literacy using a five-point Likert scale. This 21-item survey was adapted from previously validated scales (Lintner, 2024) and captured six complementary subdimensions of (i) knowing and understanding AI (e.g., *"When I am unsure about something presented by AI, I check whether it is true or not"*), (ii) using and applying AI (e.g., *"Depending on the situation, I can choose suitable AI platforms"*), (iii) evaluating and creating AI (e.g., *"I can evaluate the accuracy and reliability of AI-generated responses"*), (iv) ethical dimension of AI (e.g., *"I can identify potential ethical issues related to AI use"*), (v) social impact of AI (e.g., *"I consider the societal effects of AI advancements"*), and (vi) detecting AI (e.g., *"I can tell if I am interacting with an AI application or a human"*). Internal consistency for our sample was acceptable (Cronbach's α = 0.76).

Figure 1

Study Design and Relevant Measurements (M_x) Within Each Phase



Subsequently, the focal topic of class discussion was introduced – *Is the use of AI / genAI / LLMs justified in educational contexts?* Participants were then randomly chosen to interact with one of three custom chatbots that we had designed via backend prompts and a curated knowledge base – *TechBot*, *MarketMind*, and *EthosBot* – for up to 30 minutes. Drawing on Selwyn (2024), our rationale was to design sufficiently distinct and orthogonal perspectives, specifically in the form of a technical proponent, a corporate proponent, and a social critic of AI for education. Thus, *TechBot* was nudged to maintain an enthusiastic stance towards AI by focusing on technical details, engineering advancements, and positive outcomes of AI in educational settings. *MarketMind* was nudged to maintain a business-minded stance towards AI adoption by focusing on aligning educational outcomes with workforce needs, market demands, and corporate practices. *EthosBot* was nudged to maintain a cautious and

 © ISLS

critical stance towards AI by focusing on advocacy for the societal impacts of AI for education (e.g., social justice, equity, etc). In addition to specific instructions relevant to their conversational contexts, all three chatbots were given common guidelines to maintain their personalities, play devil's advocate when prompted to adopt views on AI that differ from their personalities, and not reveal their personality under any condition. Further, following best practices in prompting for learning (Sharples, 2023; Mollick & Mollick, 2024), all chatbots were given general instructions to vary the conversational rhythm by mixing explanations and probing questions, refrain from presenting long blocks of text and instead aim to have a good back-and-forth, and finally not do the thinking for the human partner but rather build on what they know. Some example interactions were provided as part of the backend prompt for affording extrapolation of expected response styles in the AI's conversational turns. ChatGPT 4o was used as the design platform for the present work. Authors can be contacted for complete backend prompts.

In the next phase, all participants individually completed a 20-minute pre-reflection where they answered three sets of questions. First, their *understanding of the stance of the chatbot* (M_1), collected via open-ended questions focusing on the central issue discussed with respect to the discussion topic, the chatbots' perspective on the issue, along with which of its arguments seemed most convincing in supporting that perspective. Second, their *defense of the stance of the chatbot* (M_2), collected again via open-ended questions that probed students on how they would defend the chatbot's reasoning against criticisms including strategies like refuting or downplaying opposing views and counterarguments. Third, their *initial stance on the discussion topic* (M_3), gathered via an open-ended question evaluating how has trying to adopt and defend the chatbot's perspective influenced their understanding. Participants could refer back to their chatlogs when answering these reflection questions.

Everyone subsequently moved to a groupwork phase (45 minutes). Here, we allowed teachers to form groups of three based on their seating arrangement in class, which organically led to $N = 2$ groups where everyone had previously interacted only with *TechBot*, and $N = 5$ other groups with prior interactions spanning any two ($n = 4$) or all three chatbots ($n = 1$). One group had to be excluded from all analyses due to a lack of consent. We used a worksheet to scaffold this collaborative discussion. Every group member first summarized their own chatbot's perspective briefly (10 minutes). This was followed by analyzing arguments put forth by each group member (25 minutes), specifically by articulating core claims, evidentiary basis (reasoning), underlying implicit and explicit assumptions, and counterarguments (limitations) for each line of argument within a table. We laid out norms of this discussion by indicating that everyone was expected to respectfully challenge ideas and critically analyze arguments, with the goal of exploring different perspectives on an issue rather than simply winning. Finally, the group as a whole was nudged to note down primary areas of agreement as well as disagreement (10 minutes), implicitly laying out the expectation that achieving consensus was not necessary. Participants were required to use different font colors in the worksheet to mark their individual contributions clearly.

The final phase of our study design comprised a second post-reflection activity (20 minutes), where everyone answered three sets of questions. First, *benefits of interacting with peers* (M_4), collected via open-ended questions that focused on valuable insights gained from groupwork (e.g., because of a viewpoint that challenged initial understanding) and specific ways in which their understanding evolved due to the discussion. Second, *benefits of engaging with chatbots individually prior to collaboration with peers* (M_5), gathered via an open-ended question that probed teachers for whether and how their individual AI-assisted preparations contributed to the analysis of different arguments put forth during groupwork. Third and finally, we captured everyone's *revised stance on the discussion topic* (M_6) via an open-ended question asking for elaboration of key reasons and evidence supporting their current stance post-groupwork. We ended with a debrief of the entire activity sequence.

Research questions

RQ1a: How did participants describe their understanding of a biased chatbot's stance on the learning topic?

RQ1b: How did participants describe their attempts to defend the biased chatbot's stance on the learning topic?

RQ2: What was the impact of collaborating with human peers on the evolving understanding of participants?

RQ3a: What was participants' experience of engaging in AI-assisted individual preparation prior to collaboration?

RQ3b: How was participants' own stance on the learning topic affected after interacting with AI and their peers?

Analysis plan

To answer our research questions focused on *human-AI collaboration* (RQ1), *human-human collaboration* (RQ2), and the *overall effectiveness of our learning design* (RQ3), AI-assisted thematic analysis was employed with varying qualitative data sources as inputs – M_1 for RQ1a, M_2 for RQ1b, M_4 for RQ2, M_5 for RQ3a, and a comparative analysis of M_3 and M_6 for RQ3b. Initially, generative AI was used to inductively identify themes by segmenting at the speaker-level to capture each teacher's unique perspective while preserving enough context. A novel three-part prompting framework, drawing on Sinha (2026), was used, where we started with an (i)

exploratory phase, crafting deliberate instructions for the AI to come up with a larger set of key themes in the data (e.g., 5-7) with supportive evidence and justification, followed by, (ii) a *focused refinement* phase, again crafting deliberate instructions for the AI to reconsider all themes by thinking through if they were exclusive enough, evaluate and justify the possibility of keeping a smaller more parsimonious set of themes that best explain the data, and subsequently (iii) *collaborative refinement*, where the smaller set of AI-identified themes were manually evaluated and reorganized by the lead researcher, as a content expert, for validity so as to ensure that the themes were conceptually coherent, theoretically relevant and interpretively meaningful. A codebook based on the verified themes was then created to involve two independent human raters (blind to the individual-level participant demographics), who applied themes to 20% of the data deductively. Teacher reflections sometimes contained multiple clauses, inherently allowing for double coding – thus, in the case that more than one code was applicable, following one common practice for resolving ambiguities (Chi, 1997), the more dominant code best representing the core meaning of the response was systematically picked. Inter-rater reliability, computed between the expert-validated themes and thematic coding from two-three independent raters accounting for their dominant code, was deemed acceptable for the various data sources corresponding to RQ1a, RQ1b, RQ2, RQ3a and RQ3b (Krippendorff's α = 0.81, 1, 1, 0.85, and 0.73). Disagreements were resolved by discussion, after which the remaining data was human-coded. *GPT 4o*, accessed via Azure OpenAI service, which allowed safeguarding participant privacy by making prompts and anonymized data unavailable outside the analysis context, was used.

Results

Manipulation checks

Baseline AI literacy levels of participants (M = 3.88, SD = 0.30) were deemed sufficient to engage in meaningful interactions with our designed chatbots. Conversation logs during initial chatbot interactions comprised M = 17.31 (SD = 11.52) turns, with the average conversation length per turn being M = 195.05 (SD = 116.06) words for the chatbot and M = 23.9 (SD = 20.17) words for the human conversational partner. However, despite this observable human succinctness and AI verbosity, application of a simple conversational quality metric evaluating the level of critical analysis in human turns (0 – human accepts AI feedback at face-value without question or reflecting on the AI feedback, e.g., using *ok*, *yes*, *thank you*; 1 – human occasionally questions or reflects on AI feedback without providing a new perspective, e.g., simply reflecting/responding to a question posed by AI; 2 – human regularly assesses AI feedback critically, asks for clarifications or justifications, and provides a new perspective that was not originally prompted by AI; 3 – human demonstrates deep critical analysis of AI responses by pushing for deeper explanations and seeking an understanding of the rationale behind its suggestions, e.g., by questioning AI feedback), drawing on Sinha (2025a), suggested that participants showed responsible human-AI collaboration. This was seen in the majority of our participant turns (60.6%) being coded into levels 2 and 3, as coded by two independent raters after achieving Krippendorff's $\alpha > 0.7$. Taken together, this suggests thoughtful engagement during our designed chatbot interaction activity.

Group worksheets during the later peer interaction were coded for two objective and observable measures to evaluate the fidelity of collaboration. First, *equality of participation* was quantified using the Gini coefficient, which is a widely used measure of statistical dispersion – in our context, this measure would capture how far the distribution of the number of times each group member contributed in the argumentation subphase with either a claim, reasoning, assumption or counterargument deviates from perfect equality (*min.* 0, *max.* 1). On average, the Gini coefficient for our data sample was M = 0.04 (SD = 0.06) across the N = 7 groups, indicating high *equality of participation*. Second, we also looked at evidence for *transactivity* during the argumentation subphase by noting down instances in group worksheets where multiple font colors (proxies for different individual contributions) were used as part of (i) either the same core argument thread (e.g., if member A stated the claim, member B stated the assumption for that claim, and member C stated the counterargument for the same claim, three transactive contributions would be counted), or, (ii) the same claim, reasoning, counterargument, or assumption (e.g., if member A and member B both contributed to fleshing out a particular claim, one transactive contribution would be counted). We found that the presence of this higher-order collaborative skill was rarer, at least when we made inferences only from the group worksheets, as only two of the N = 7 groups showed evidence of building on each other's ideas rather than merely moving their own arguments and perspectives forward. Memos from our in-class observations, though, did suggest participants to be quite active during this argumentation sub-phase, and we posit that time constraints may have prevented *transactivity* to be accurately reflected in the group worksheets.

Understanding of the chatbot's stance (RQ1a)

Our thematic analysis results for RQ1a suggested three core themes for participants' foundational understanding of the chatbot behaviors. First, *recognition of the chatbot bias*, defined as reflecting on how directional biases are

presented in the chatbot's stance either in terms of ethical evidence, economic logic, or techno-optimism, and exemplified via quotes like (i) "Actually at first it felt very efficiency and effective stance but if I would say something which is most convincing would be the Bots perspective being the 'Evidence-Based Efficacy' claim – specifically that ITS and AI writing assistants show measurable improvement in student learning outcomes", (ii) "I don't find the arguments strong because the AI cannot understand what my subject discipline and work requires. It keeps coming from the economic perspective and doesn't take into account other perspectives like ethics and assessment of students", and (iii) "AI kept trying to monetize the idea, marrying that to the efficacy of the product…starting small at a classroom level, because it is lowest cost (effort), highest learning environment".

Second, *recognition of the conditions for making AI a valuable tool*, defined as making judgments on how AI is used and whether it is anchored with an intentional pedagogical purpose, and as seen in quotes like (i) "However, when AI is doing more – teacher's role will slowly 'disappear' and the child might just want to communicate with AI instead of having interaction with human...AI can become a trusted companion…But only teacher can help children translate that trust into real-world relationships", (ii) "Over-reliance can erode thinking, creativity, and responsibility...AI should support, not replace, human thinking in education...When designed intentionally, they can prompt critical thinking and clarify complex ideas, helping both students and teachers engage more meaningfully with learning", and (iii) "The teacher must and can adapt alongside with AI...The teacher becomes an experience creator – making use of AI to further supplement areas which may be limited by human ability (e.g., personalization or instant feedback to every student)".

Third and finally, *recognition of cultural and contextual specificity for evaluating AI's role*, defined as attempting to question how localization, values and national identity may conflict with AI's embedded norms, and exemplified via verbalizations like (i) "Quoting the chatbot: I believe Singapore absolutely needs to invest in building its own AI tools for education – but not just for the sake of being self-sufficient. It's about protecting our values, students, and teachers from being shaped by someone else's idea of what learning should be" and (ii) "To ensure ethical and pedagogical alignment, educators, students, parents, and education authorities must all be involved in shaping how AI is designed and used in schools. We must prioritize tools that serve learning and inclusivity – not just efficiency".

Defense of the chatbot's stance (RQ1b)

Our thematic analysis results for RQ1b suggested three categories grounded in participants' descriptions of how they attempted to defend the chatbot's stance. First, *defending through reframing*, defined as casting the chatbot's advocacy for AI adoption as a means to enhance professionalism, equity and teaching practice, exemplified via quotes like (i) "The bot's argument is strategically powerful…bypasses ideological resistance by focusing on outcomes rather than process…appeals to professional accountability…provides institutional justification for resource allocation and policy decisions", (ii) "I am convinced by its [chatbot's] ability to support learners across different ranges. It is able to present great ideas on how AI is able to support the needs…", and (iii) "Someone might say that AI might deskill teachers or reduce their professional autonomy. I can defend the chatbot by advising the person that AI can empower teachers by removing routine tasks, freeing time and mental energy to focus on higher-level professional work like mentoring, creativity, and individualized student support. Far from diminishing autonomy, it can enhance teachers' roles by giving them more space to exercise their expertise".

Second, *defending through dialogic balance*, defined as justifying the chatbot's bias via meta-level discourse features such as its responsive tone (e.g., open, balanced) and argumentation style, exemplified via quotes like (i) "It often carefully considers the argument rather than outrightly rejecting it. It will take elements from the counter argument…and use them to further develop on its main argument", (ii) "The chatbot would still consider the perspectives of the person and provide a well-rounded argument. Having done so, it would still leave the final decision to the inquirer", and (iii) "It didn't downplay it, in my opinion. It gave me alternative viewpoints to consider without sounding too insistent that I should accept its stance".

Third, and finally, *defending through ethical imperatives*, defined as pivoting to ethical imperatives by positioning AI literacy and human judgment as social/moral necessities to protect learners, as seen in quotes like (i) "But if there was someone who hold this view that teachers are already too overwhelmed to also teach students digital literacy, I would defend the chatbot and say that AI is going to be ever pervasive – if we don't teach our students AI literacy and how to interrogate bias, they would be at risk of accepting things at face value and not being able to think and problem solve for themselves. Then whoever is controlling AI will control their minds" and (ii) "If someone argued that AI could fully replace human discipline and ethics by perfectly monitoring and correcting behavior, it cannot instill genuine values or emotional resilience. True accountability and ethical judgment come from lived human experience and internal motivation, which AI cannot replicate. Relying solely on AI risks creating blind trust in technology, ignoring the complex, nuanced decisions humans must make, especially in unpredictable, high-stakes environments where empathy and moral reasoning are essential".

Perceived benefits of human-human collaboration (RQ2)

Participants' descriptive evaluation of the perceived benefits of interacting with their classmates fell into two broad thematic categories. Our analyses focused on participants' explicit articulations about how their prior beliefs expanded via either a change or reaffirmation due to groupwork. First, *change* was evident in quotes like (i) "It changed a little as I acknowledged that educational context does not necessarily mean only for students but it can be used to lighten teachers' admin load and prep work. My group members highlighting how if we manage to get more teachers on board to use AI in lessons and to teach AI literacy, then there will be more ROI [return on investment]. So, this might make the use of AI in education more justifiable as the initial concern was teachers being overworked to take on the task of ensuring that AI is equitable and promoting critical thinking, but if more teachers get on board, workload can be split ensuring ROI", and (ii) "The most valuable insight I gained from hearing the perspectives of others in the group was how diverse and creative the use of AI can be across different teaching contexts. It was eye-opening to see how each educator tailors AI tools to address the unique needs of their students and subject areas. Such as: How AI was used to support students with special needs through personalized content, in PE to provide real-time feedback on movement".

Second, *reaffirmation* of beliefs was exemplified via quotes like (i) "I believe that it has evolved as we delve deeper into the discussion. AI is definitely here to stay and there will be many more generations of AI to come. While we do not need to embrace it with open arms, we can be discerning and make our choices. Learn about how to detect bias in AI responses using AI platforms (a bit ironic!) and make informed choices. I feel that the group discussion was good as our responses are rather similar as it affirms our common beliefs that AI is meant to augment teachers' lessons and not replace them", and (ii) "It did not change, as my group actually do agree that while AI can improve efficiency and productivity, the ethical part of using AI in educational context is still unjustified. The arguments to support why use of AI is ethical is merely based on the fact that it provides diverse perspectives. But it is too narrow an argument to justify why AI use in educational context is justified, relative to the costs of creating learner passivity. We can't be sure that every student will engage with AI in an ethical manner that builds critical thinking and gives exposure…".

Engaging in the designed groupwork also led to critical reflections on *implications of responsible AI use* for education, as reflected in one participant's abstraction – "… it [groupwork] provided me with new ways to engage teachers who are at the extreme ends of practice – the Pencil Metaphor – those at the tip, and those at the eraser end. It also gave me ways to engage school leaders who enforce a blanket rule on the use of AI, disregarding the nature of the subject, its pedagogical uniqueness and students experiences. It supports my initial concerns, and provided me with a wider perspective to still consider it for use by PE [physical education] teachers".

Perceived benefits of individual preparation before collaboration (RQ3a)

When asked to share perceived benefits, if any, of engaging in our learning design, participant responses fell into three themes. First, *cognitive and reflective growth*, defined as judgments of how chatbot preparation helped clarify thinking, understand multiple perspectives and led to individual stance formation on the learning topic, and exemplified via quotes like (i) "Preparing with the chatbot really helped me understand the topic better. It explained both sides of the argument clearly and gave me useful points I could bring into the group discussion…What helped most was realizing how my own view changed – I started off unsure…but learning how interactive features can support those with ASD or selective mutism gave me a new perspective", and (ii) "I think in essence I already had my stance and the bot even though tried to sway me to a Yes also agreed in a number of my concerns which was also further agreed by my teammates".

Second, *collaborative growth*, defined as judgments for how the group was collectively supported in analyzing arguments, surfacing assumptions, or having individual members contribute unique views drawn from their specific chatbot's stance, and as seen in quotes like (i) "The ideas and examples provided by the chatbots provides support to counter arguments with regard to the negative perspective of AI in education. It allows group members to see how AI is able to support diverse learners, but this depends on the correct pedagogical design and strategy implemented by the teacher", (ii) "Different individuals have different areas that they want to focus on when discussing the chatbot on 'Is the use of AI / genAI / LLMs justified in educational contexts?'. So, individual preparation helps in bringing up points to share with my team mates…", (iii) "Most of the ideas presented during the group discussion just reinforced the ideas brought up by the bot. I think that the human interaction and cross pollination of ideas then resulted in new ideas formed", and (iv) "[chatbot preparation] provided more in depth discussions as some points may be missed without inputs from the chatbot".

However, we also had a third thematic category capturing *mixed and critical appraisals* of the chatbot's usefulness as preparation for groupwork, exemplified via skeptic verbalizations such as "…it [chatbot] does not really contribute to the analysis of the different arguments presented during the group discussion. It [group] is just

a sharing of the ability and limitations of AI". Two of the n = 20 participants explicitly stated that they did not understand the AI-assisted preparation or did not find it contributing directly to groupwork.

Evidence for the evolution of learning (RQ3b)

Our comparative analyses of participants' *initial stance* after their individual chatbot interaction (M_3) and *revised stance* post-groupwork (M_6) suggested three parsimonious thematic accounts of how stances shifted (or didn't). The first most frequent thrust (60%) was representative of a *cautiously positive adoption (endorsement)* that kept teachers central and justified AI use only when it augmented teachers and helped diverse learners rather than replacing pedagogy. In other words, the quotes in this first thematic category advocated for a conditional, teacher-partnered adoption of AI grounded in student-centered benefits. The second relatively rarer thematic category (20%) comprised a shift towards (or reaffirmation of) justifications centered on *ethical and epistemic guardrails*. Because this theme emphasized AI guardrails that participants demanded (e.g., bias, monitoring, AI literacy or epistemic validity), coders evaluated whether participants foregrounded normative conditions for responsible AI usage as opposed to teacher-partnership or personalization claims in their responses. The third and final theme (20%) centered on *pragmatic constraints* such as feasibility, workload or day-to-day-practicality (e.g., time, data upkeep, who will build/maintain AI etc) in participants' responses. Simply put, distinct from positive adoption or ethical doubts, this theme foregrounded operational constraints of AI usage. See Table 1 for an overview.

Table 1

Comparative Analyses of Participants' Initial and Revised Stance on The Learning Topic

Theme	Stance	Exemplar quotes from the data
#1	Initial	"I have somewhat developed a deeper appreciation for the practical balance between AI's strengths and its limitations in education. Overall, defending the chatbot's view somewhat shifted my understanding from seeing AI as either the panacea or a silver bullet or avoid-it-at-all-costs towards a more nuanced, balanced perspective. AI is very real and here to stay but very much is dependent on thoughtful, intentional human partnerships with AI".
	Revised	"AI may have manifold advantages in supporting teachers in the classroom. However, it lacks the personal touch that teachers have. AI can suggest the first draft in a lesson, the teacher will curate and refine before delivering the lesson. This is because there are students with different levels of readiness and learning needs. These are critical information that AI doesn't have. During my chatbot interaction, it was quick to point out the benefits of AI as well as acknowledge that it must work together with the teacher to achieve educational outcomes in the classroom. During my group discussion, we were a bit disheartened when we realize that AI is taking over many jobs e.g., healthcare industry. However, when we think deeper into the issue, we realize that AI can't take over 100%! For instance, AI vs human doctor in informing a patient that she has an aggressive brain cancer and has only 3 - 6 months left to live".
#2	Initial	"In my discussion with the chatbot, I already hold the same view. It is interesting because prior to this, I never really thought about how training data and design assumptions might influence my students learning".
	Revised	"My current stance remains the same and the chatbot serves to support my stance that the use of AI/ genAI / LLMs is only justified in educational contexts when it promotes critical thinking. At the moment, commercial AI tools aren't designed with pedagogy, equity, or student development in mind. And since they're built for scale, they are not designed to nurture nuanced thoughts. I think that this is a problem and until this can be circumvented through the teaching of AI literacy, AI use in educational contexts is not fully justified".
#3	Initial	"…because it [AI] does not have a solution the real world problem that we face with students every day. And it takes every day and every single available learning opportunity available to teach students about using AI responsibility and verifying what it says etc. It is everyday work, not a 1 or 2 lesson job".
	Revised	"Stance remain the same. No change, group discussion is in line with what I already know. response from chatbot is also within expectation".

Discussion and conclusion

We implemented a novel pedagogical approach in authentic higher education classrooms where interaction with biased AI chatbots was designed to prime teachers for engaging in groupwork with their peers. As a *learning design* contribution, we provide a pedagogical framework to bring together the complementary affordances of *human-AI collaboration* and *human-human collaboration*, while offering targeted reflection and critiquing

opportunities to optimize the learning potential from each phase. As a *theoretical* contribution, we highlight the mechanism of designed *cognitive conflict* (instantiated in the clash between the biased AI stance and a peer's counter-bias) and *perspective-taking* (analysis of multiple flawed viewpoints) to intellectually challenge learners.

Results for RQ1a, which focused on *foundational understanding of the chatbots' perspectives*, showed that teachers recognized the embedded bias we had deliberately designed for, sometimes finding it compelling, but often challenging their chatbots' narrow framing or assumptions. Teachers further endorsed a collaborative, augmentative role for AI – where AI acts as a tool, co-teacher, or assistant, and teachers retain final authority. They used an application-oriented lens to judge whether AI's use is justified and saw pedagogical intentionality and human oversight as essential to making AI use effective. Teachers also demonstrated a nuanced recognition that AI is never value-neutral and culturally-aligned design and governance are key for justified implementation in their context. Beyond simply recognizing the bias, though, we also used reflection questions that pushed teachers to consider various ways of defending the stance of their biased chatbots and making it appear credible. Our rationale was to trigger critical evaluation, perspective-taking, and deeper processing of how the chatbot's bias manifests, and have teachers focus on elaborating a strong defense for that specific viewpoint (by finding evidence and logical-sounding justifications). Results for RQ1b, which focused on *supporting the chatbot's stance while likely filtering out and/or downplaying contradictory information*, showed that teachers conceived their chatbot to be defendable when its advocacy on AI adoption could be reframed as professionally responsible, pedagogically beneficial, or operationally pragmatic. Simply put, their chatbot's stance on AI as a tool (rather than a threat) could be defended when foregrounding how AI appeals to institutional values or classroom impact. In addition to *what* the chatbot argued, teachers also highlighted *how* its rhetorical strategy (e.g., responsiveness and openness) often rationalized or softened the bias put forth. Teachers further focused on normative (value-based) justifications for the chatbot's stance based on future-oriented and societal concerns, invoking ethical obligations for preparing students for bias in AI, ensuring cultural or cognitive independence, and protecting moral agency. In summary, these analyses for RQ1a and RQ1b, beyond serving as extended manipulation checks, suggested that our probing questions (administered through measures M_1 and M_2) successfully problematized participants' understanding. This was crucial to align their personal critical reflections with the biased perspectives of their individual chatbots before beginning groupwork. Maximizing this divergence of opinions was intentional.

Moving on to evaluate the *impact of groupwork on teachers' evolving understanding* (RQ2), we found in post-groupwork verbalizations (measure M_4) that despite not achieving consensus, peer discussions may still have played an important role in sharpening or reinforcing ethical and human-centered concerns inherent in the learning topic, causing epistemic movement in (or, confirmation to) teachers' prior beliefs. When teachers were forced to reconcile information from their peers, this presumably moved them beyond simply understanding their own assigned bias to comprehending how different biases can lead to conflicting and equally flawed conclusions. Groupwork thus afforded teachers the opportunity to deconstruct this initial bias as they confronted their peers with similarly well-articulated defenses for opposing, biased positions. When evaluating their *experiences engaging with our learning design* of individual work as preparation for collaboration (RQ3a), teachers articulated that they used the chatbot as a learning partner, sometimes transformative and sometimes confirmatory, but always in relation to their own conceptual stance on the learning topic (measure M_5). They elaborated on the dialogic impact of individual preparation too (e.g., how ideas from the chatbot were operationalized into group talk). Few teachers, however, expressed skepticism towards the learning design. Finally, when looking at *how everyone's stance on the topic was affected* (RQ3b) after working with the chatbot (measure M_3) and subsequently after integrating learning from the groupwork (measure M_6), we found that for the majority of teachers, their revised positions articulated a more balanced, conditional, and criteria-driven stance relative to earlier, less-specified remarks that may have been driven primarily by interactions with our biased chatbots. The majority moved from tentative or tool-focused takes to qualified endorsement of AI with boundaries, and surfaced principled ethical criteria as prerequisites for AI adoption. Given the moderate-to-high *AI literacy* in our sample, we interpret these enhancements as *epistemically enabled* rather purely chatbot exposure-driven. However, for a minority, there still remained a tendency for (partially) rejecting change in stances on the learning topic based on feasibility grounds.

For educators, an important implication stemming from the present work is to not give learners AI to find answers, but use AI's inherent flaws (e.g., their tendency towards bias) to our instructional advantage (e.g., prime conflict / start a debate) in the service of building critical thinking skills and discouraging undue dependence on technology (Sinha, 2025b). Our limitations stem from working with a small teacher sample with relatively high levels of AI literacy. We also focused on an engaging learning topic of *AI adoption in higher education.* Future work across other topics is therefore warranted to assess the generalizability and depth of reflections engendered by our proposed learning design. Greater sample size will also afford running quantitative analyses of qualitative data (Chi, 1997) to assess correlations with short-term and long-term learning measures.

References

Asterhan, C. S., & Schwarz, B. B. (2016). Argumentation for learning: Well-trodden paths and unexplored territories. *Educational Psychologist*, *51*(2), 164-187.

Chen, W., Lyu, Q., & Su, J. (2024). The role of individual preparation before collaboration: An exploratory study on students' computer-supported collaborative argumentation in a primary classroom. *Journal of the Learning Sciences*, *33*(4-5), 757-798.

Chi, M. T. (1997). Quantifying qualitative analyses of verbal data: A practical guide. *Journal of the Learning Sciences*, *6*(3), 271-315.

Chinn, C. A., & Clark, D. B. (2013). Learning through collaborative argumentation. In *International Handbook of Collaborative Learning* (pp. 314-332). Routledge.

Dillenbourg, P. (1999). What do you mean by collaborative learning?. *Collaborative-learning: Cognitive and Computational Approaches.*, 1-19.

Kim, M., & Lim, C. (2025). AI chatbot design principles to enhance the collective efficacy in collaborative learning. *Asia Pacific Education Review*, 1-14.

Kovari, A. (2025). A systematic review of AI-powered collaborative learning in higher education: Trends and outcomes from the last decade. *Social Sciences & Humanities Open*, *11*, 101335.

Lam, R., & Kapur, M. (2018). Preparation for future collaboration: Cognitively preparing for learning from collaboration. *The Journal of Experimental Education*, *86*(4), 546-559.

Lintner, T. (2024). A systematic review of AI literacy scales. *Npj Science of Learning*, *9*(1), 50.

Mende, S., Proske, A., & Narciss, S. (2021). Individual preparation for collaborative learning: Systematic review and synthesis. *Educational Psychologist*, *56*(1), 29-53.

Mollick, E., & Mollick, L. (2024). Instructors as innovators: A future-focused approach to new AI learning opportunities, with prompts. *arXiv preprint arXiv:2407.05181*.

Nokes-Malach, T. J., Richey, J. E., & Gadgil, S. (2015). When is it better to learn together? Insights from research on collaborative learning. *Educational Psychology Review*, *27*(4), 645-656.

Rajaram, K. (2021). *Evidence-based teaching for the 21st century classroom and beyond*. Springer Singapore.

Selwyn, N. (2024). Constructive criticism? Working with (rather than against) the AIED back-lash. *International Journal of Artificial Intelligence in Education*, *34*(1), 84-91

Sharples, M. (2023). Towards social generative AI for education: theory, practices and ethics. *Learning: Research and Practice*, *9*(2), 159-167.

Sinha, T., Kapur, M., West, R., Catasta, M., Hauswirth, M., & Trninic, D. (2021). Differential benefits of explicit failure-driven and success-driven scaffolding in problem-solving prior to instruction. *Journal of Educational Psychology*, *113*(3), 530-555.

Sinha, T., & Kapur, M. (2021). Robust effects of the efficacy of explicit failure-driven scaffolding in problem-solving prior to instruction: A replication and extension. *Learning and Instruction*, *75*, 101488.

Sinha, T. (2025a). Evaluating efficacy of problematizing chatbot in higher education. In *Proceedings of the 19th International Conference of the Learning Sciences* (pp. 2629-2631).

Sinha, T. (2025b). Beyond good AI: The need for sound learning theories in AIED. *Technology, Knowledge and Learning*, *30*(4), 2443-2457.

Sinha, T. (2026). Making failure desired during learning – A quasi-experimental study. *Thinking Skills and Creativity*. 60, 102094

Thorgeirsson, S., Sinha, T., Friedrich, F., & Su, Z. (2022). Does deliberately failing improve learning in introductory computer science?. In *European Conference on Technology Enhanced Learning* (pp. 608-614). Cham: Springer International Publishing.

Vogel, F., Wecker, C., Kollar, I., & Fischer, F. (2017). Socio-cognitive scaffolding with computer-supported collaboration scripts: A meta-analysis. *Educational Psychology Review*, *29*(3), 477-511.

Weng, X., Xia, Q., Gu, M., Rajaram, K., & Chiu, T. K. (2024). Assessment and learning outcomes for generative AI in higher education: A scoping review on current research status and trends. *Australasian Journal of Educational Technology*, *40*(6), 37-55.

Zhang, R., Zou, D., & Cheng, G. (2024). A review of chatbot-assisted learning: pedagogical approaches, implementations, factors leading to effectiveness, theories, and future directions. *Interactive Learning Environments*, *32*(8), 4529-4557.

Acknowledgments

Authors are supported by Social Sciences and Humanities Research grant from Nanyang Technological University (2024-CoHASS-023). Thanks to Shivaangii Jaiswal, Koek Wei Jie Dominic, and Shen Yushu for their assistance.

International Society of the Learning Sciences

How Simulation Design Shapes Academically Productive Talk: An Epistemic Network Analysis of Teacher Discourse in Middle-School Physics

Wei Jia, Kexin Li, and Gaowei Chen
wjia627@connect.hku.hk, kxmichelle@connect.hku.hk, gwchen@hku.hk
The University of Hong Kong

Abstract: Dialogic teaching, through Academically Productive Talk (APT), enhances student learning. While digital simulations are common in science education, their specific impact on teacher discourse patterns remains underexplored. This study examines how simulation design, Multiple Representations (MR) versus Single Representation (SR), shapes whole-class discourse structure and teachers' use of APT moves. One teacher and 99 eighth-grade students from two classes in Western China participated. Using Epistemic Network Analysis (ENA) and discourse analysis of 147 teacher talk turns, we compared APT patterns between two conditions. ENA revealed distinct strategic repertoires: the MR condition showed an integrated, collaborative network centered on reasoning and building on ideas, involving more students in shorter, focused turns. The SR condition exhibited a linear, teacher-guided pattern dominated by elicitation and challenge, with longer explanations from fewer students. Overall, simulation design relates to teacher facilitation and student participation, suggesting that multiple representational tools may afford more collaborative-dialogic classrooms.

Introduction

Dialogic teaching enhances student engagement and learning outcomes through structured teacher-led discourse (Michaels & O'Connor, 2015; Tao & Chen, 2024). With professional development, teachers can employ strategies such as Academically Productive Talk (APT) to foster more active and cognitively demanding classroom environments, encouraging deeper thinking and collaborative knowledge building (Bouton et al., 2025; Chen, Chan, et al., 2020). In parallel, digital learning environments, particularly computer simulations, have gained prominence for their interactive and visual affordances, which support conceptual understanding in science education (Li et al., 2024; Testa & Catena, 2022). The design of simulations, including their representational complexity, significantly influences learning processes and outcomes (Jaakkola & Veermans, 2020; Trypke et al., 2024). However, extant research on APT analyzes dialogue at an aggregate classroom level or focuses narrowly on teacher behaviors, potentially overlooking nuanced interaction patterns shaped by specific instructional tools. This study investigates how different simulation types, namely multiple versus single representations, shape whole-class discourse structure and teachers' APT patterns in junior high physics classrooms.

Academically productive talk

Dialogic pedagogy encompasses several frameworks aimed at fostering productive classroom discourse, including Academically Productive Discussion (Michaels & O'Connor, 2015), Accountable Talk (Michaels et al., 2008), and Productive Classroom Dialogue (Pehmer et al., 2015). This study adopts the framework of Academically Productive Talk (Resnick et al., 2010), which is grounded in Vygotsky's (1978) sociocultural theory and situated learning perspectives (Brown et al., 1989). APT emphasizes knowledge co-construction through social interaction, as students verbalize, share, and refine their thinking collectively (Michaels & O'Connor, 2015). Table 1 outlines eight APT talk moves aligned with four instructional goals, contextualized for physics learning (Chen, Chan, et al., 2020).

Research indicates that productive discourse positively influences student engagement (e.g., emotionally, cognitively, and behaviorally), leading to improved learning outcomes (Sedova & Sedlacek, 2023; Tao & Chen, 2024). Teachers' adept use of APT moves is crucial for fostering such discursive engagement, a characteristic of high-quality instruction (Bouton et al., 2025; Chen, Zhang, et al., 2020). Consequently, recent efforts have emphasized teacher professional development (TPD) aimed at integrating productive talk to promote more active and equitable student participation (Bouton et al., 2025; Wang, Yang, et al., 2025). Nonetheless, few studies have investigated how digital learning environments, such as simulations, shape teachers' APT moves and associated discourse structures in physics education.

Table 1
Teachers' APT Moves Adapted from Previous Studies (Chen, Chan, et al., 2020; Resnick et al., 2010)

 © ISLS

Goal	APT Moves	Description	Examples
Elaborating	Say More	Prompt students to elaborate on their initial ideas	"S1, can you elaborate on what this animation is demonstrating?"
	Revoice	Revoice student reasoning and allow them to confirm its accuracy	"S1, are you suggesting that this animation depicts the melting process of a crystalline substance?"
Reasoning	Press for Reasoning	Ask students to articulate the reasoning behind their ideas	"S1, why do you interpret this animation as representing the melting of a crystalline substance?"
	Challenge	Challenge students' thinking or present a counterexample	"S1, if this were depicting the melting of an amorphous substance, such as asphalt, what would the animation look like instead?"
Listening	Restate	Request that students restate or paraphrase a peer's ideas	"Can another student—S2, please restate S1's interpretation of this animation?"
Thinking with Others	Agree & Disagree	Ask students to apply their own logic to a classmate's argument	"Does another student—S2 agree with S1's description regarding the melting process shown in the animation? (And why?)"
	Add On	Encourage students to build upon the ideas of others	"Can another student—S2 build upon or add to S1's explanation?"
	Explain Other	Prompt students to explain a peer's line of thinking	"Can another student—S2 explain the reasoning behind S1's interpretation?"

Multiple representations in simulations

Multiple representations (MR) employ diverse external formats, such as graphs, animations, and tables, to connect abstract scientific concepts with real-world phenomena (Ainsworth, 2006). In contrast, single representation (SR) approaches use one modality, offering limited support for comprehension (Jaakkola & Veermans, 2020). Integrating MR into dynamic simulations has shown strong efficacy in science learning (Li et al., 2024; Testa & Catena, 2022). Empirical studies indicate that MR simulations support conceptual transfer and long-term retention by facilitating transitions between representations (Ainsworth, 2006; Rexigel et al., 2024). Cognitively, they reduce working memory load and manage cognitive demands more effectively (Schwamborn et al., 2011; Trypke et al., 2024). Motivationally, vivid visual elements enhance engagement (Magner et al., 2016), while the multi-perspective nature of MR fosters deeper reasoning and supports personalized learning (Almasri, 2022). Despite these benefits, few studies have examined how different representational designs influence the quality of classroom discourse, particularly the use of APT in physics instruction.

The present study

This study examines how simulation representational design (MR vs. SR) is associated with the enactment of productive classroom discourse. Grounded in Bandura's (1986) social cognitive theory, which posits reciprocal interactions among environment, behavior, and personal factors, we treat representational design (MR vs. SR) as a key environmental variable. We expect that MR simulations, by providing diverse semiotic resources, may be associated with more interconnected APT strategies (e.g., integrating reasoning prompts with collaborative additions). This strategy pattern may co-occur with more distributed student participation, consistent with the socio-constructivist orientation of APT. To examine these dynamics, we employ Epistemic Network Analysis (ENA) to model the co-occurrence structure of teacher APT moves, moving beyond frequency counts to capture strategic patterns. Our research addresses the following research questions (RQ):

- RQ1. How do MR and SR simulation environments differ in the overall structure of classroom discourse?
- RQ2. What are the differences in the network structure of teacher APT moves between the MR and SR conditions?
- RQ3. How are these quantitative differences manifested in the qualitative flow of classroom dialogue?

Method

This study utilized data from a larger intervention program that integrated simulation-based instruction to foster productive talk in junior high school physics classrooms in Western China. The program was implemented during the fall semester of 2024 after obtaining formal ethical approval.

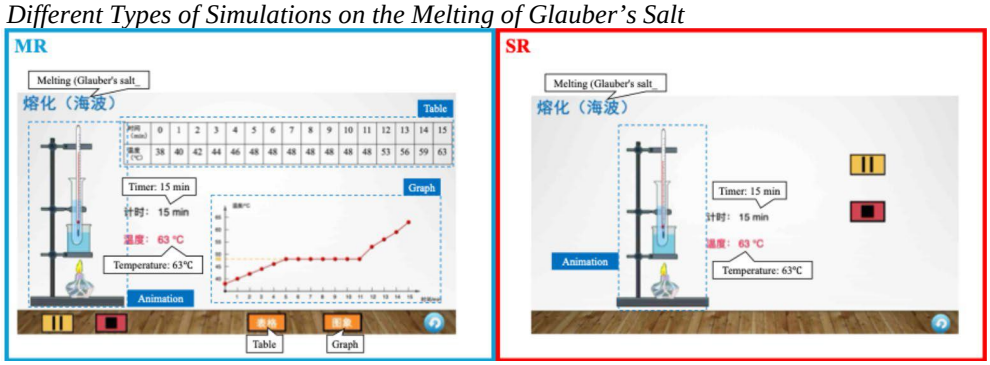
Participants and procedure

The participants included one male physics teacher (who held a master's degree and had over one year of teaching experience) and 99 eighth-grade students from his two parallel classes. The teacher voluntarily took part in the intervention and received three 40-minute professional development workshops on employing APT and simulations in instruction.

The two classes, comparable in size and overall academic performance, were randomly assigned to Class A (n = 47; 22 males, 25 females) and Class B (n = 52; 28 males, 24 females). The instructional topic, *Melting and Solidification*, was selected based on consultations with five experienced educators. The teacher utilized two types of pre-made simulations (e.g., depicting Glauber's salt and ice melting; Figure 1) tailored to the local curriculum to ensure ecological validity. In Class A (the Multiple Representations, or MR, group), students engaged with simulations featuring toggleable representations (animation, table, graph, formula). In Class B (the Single Representation, or SR, group), simulations displayed only one representation at a time. In this study, a representation refers to a domain-specific external depiction of the target phenomenon that uses a distinct semiotic system (e.g., dynamic animation of states/particles, a numerical data table, a Cartesian graph). On-screen labels or brief instructions were treated as interface text rather than additional disciplinary representations. Classroom sessions were video-recorded and subsequently transcribed verbatim, with speakers identified in consultation with the teacher.

Figure 1

Different Types of Simulations on the Melting of Glauber's Salt



Coding classroom dialogue transcripts

Transcripts were segmented into talk turns, defined as uninterrupted stretches of talk by a single speaker and bounded by a changing speaker. A total of 292 talk turns (MR: 150; SR: 142) were transcribed. Teacher talk turns (147 total; MR: 75; SR: 72) were coded using an established APT framework (Chen, Chan, et al., 2020; Resnick et al., 2010), which includes the categories: Say More, Revoice, Press for Reasoning, Challenge, Restate, Agree & Disagree, Add On, and Explain Other (Table 1). Each teacher talk turn could receive multiple APT codes when multiple moves occurred. Teacher turns that did not enact any APT move (e.g., routine information-seeking questions or conventional topic initiations) were coded as "None". Inter-rater reliability was established by two researchers performing independent cross-coding. Cohen's kappa was calculated on a code-by-code basis, revealing substantial to almost perfect agreement with values ranging from .705 to .926 (M = .81; Landis & Koch, 1977).

Method of analysis

Classroom Discourse Analyzer

To examine the overall structure of classroom discourse (RQ1), we utilized the Classroom Discourse Analyzer (CDA; Chen, 2020; Wang, Chen, et al., 2025) to automatically analyze transcripts for low-inference features, including the number of student speakers, word counts per turn, and teacher-student turn-taking patterns. Furthermore, high-inference discourse features, such as the specific characteristics of utterances, were identified and coded manually.

Epistemic Network Analysis

To address RQ2, we used Epistemic Network Analysis (ENA) to model the co-occurrence structure of APT codes in teacher talk (Bowman et al., 2021). ENA is useful for examining not only how often discourse moves occur but also how they are connected within segments of interaction. Analyses were conducted using the ENA Web Tool (v. 1.7.0; Marquart et al., 2018). We applied a moving stanza window of 9 lines to model links between current and prior codes, balancing variance explained on SVD1 and model stability. Group differences in network

 © ISLS

structure were tested using Mann-Whitney *U* tests on the resulting projection coordinates. ENA visualizations consist of points in a low-dimensional space and corresponding network graphs, where node positions and connection strengths illustrate the structure of code associations. We compared the resulting network structures across the MR and SR conditions and interpreted them together with qualitative transcript analysis for RQ3.

Results

This section reports the results of the comparative analysis of dialogic teaching in the Multiple Representations (MR) and Single Representation (SR) conditions. We first outline the overall structure of classroom discourse, then detail the patterns and network structure of the teacher's APT moves, and finally contextualize these quantitative findings with illustrative dialogue excerpts.

Comparison of overall structure of classroom discourse

While the proportion of teacher talk was comparable across groups (MR: 50.0%; SR: 50.7%; Table 2), the structure of student participation differed markedly. The MR condition showed a higher proportion of individual student turns (35.3% vs. 23.9%) and engaged nearly twice as many unique speakers (22 vs. 12). Conversely, the SR condition was characterized by a higher proportion of choral responses (25.4% vs. 14.7%). Differences also emerged in contribution length: on average, student turns were shorter in MR than in SR (8 vs. 16 words per student turn), whereas the average total contribution per student speaker was higher in SR than in MR (45 vs. 19 words per student speaker).

Table 2

Comparison of Teacher and Student Talk in Whole-Class Discourse Between MR and SR Conditions

	MR (n = 47)	SR (n = 52)
Teacher turns (T)	75 (50.0%)	72 (50.7%)
Student turns (S)	53 (35.3%)	34 (23.9%)
Choral response turns (Ss)	22 (14.7%)	36 (25.4%)
Number of student speakers	22 (46.8%)	12 (23.1%)
Mean words per student turn[a]	8	16
Turns per student speaker	2.36	2.83
Total words per student speaker (Mean) [a]	19	45

Note. [a] Words are equivalent to Chinese characters

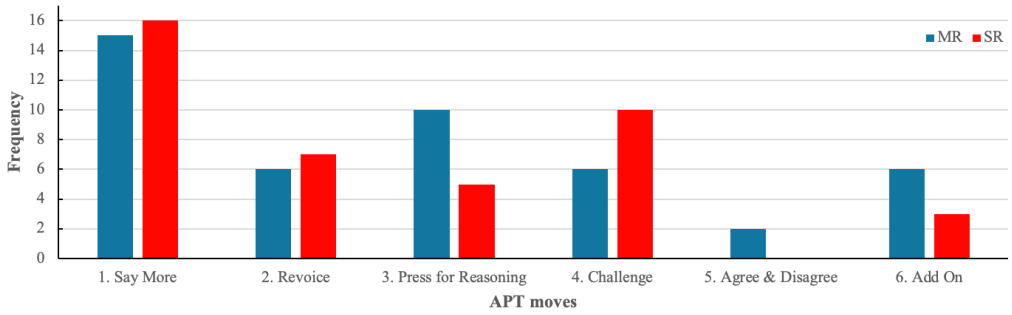
Patterns and structure of teacher APT moves

Frequency of teacher APT moves

Analysis focused on six observed APT moves, excluding Restate and Explain Other. These moves constituted 55.6% of teacher talk in the SR group and 52.0% in the MR group, calculated as the number of teacher turns containing at least one of the six APT moves divided by the total number of teacher turns in each condition (SR: 72; MR: 75). The MR condition was characterized by more frequent use of Press for Reasoning and Add On and was the sole context where Agree & Disagree was observed, albeit rarely. The SR condition featured a predominant use of Challenge and Say More (Figure 2).

Figure 2

Frequency of APT Moves in MR and SR Conditions



Network analysis of teacher APT moves

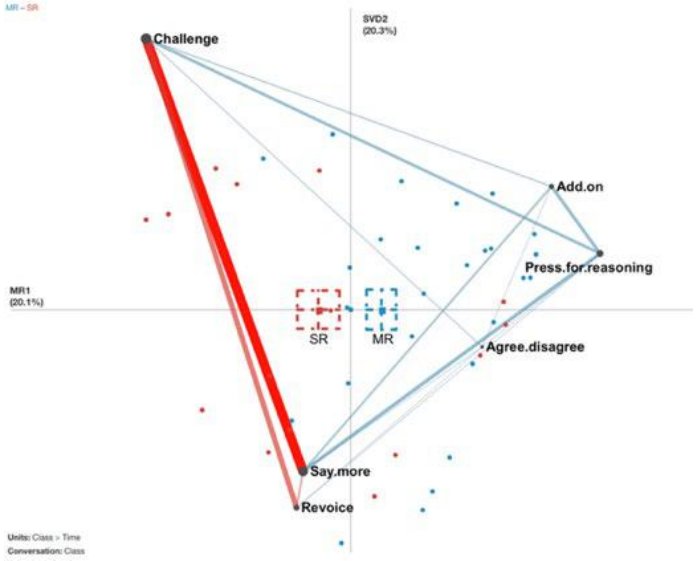
ENA indicated distinct strategic repertoires between the two groups (Figure 3). The model demonstrated excellent goodness-of-fit (x-axis: Pearson r = .96, Spearman ρ = .94; y-axis: r = .97, ρ = .97). The primary dimension

 © ISLS

(SVD1/MR1, *x*-axis), accounting for 20.1% of the variance, effectively separated the conditions. The SR group's centroid (left) was associated with Say More and Revoice, suggesting a pattern of eliciting and clarifying. The MR group's centroid (right) was linked to Press for Reasoning and Add On, indicating a pattern of probing reasoning and building collaboration. A Mann-Whitney *U* test confirmed significant separation along this dimension ($U = 1831.00$, $p < .001$, effect size $r = .33$). No significant difference was found on the secondary dimension (SVD2, $U = 2589.00$, $p = .55$, $r = .05$).

Figure 3

Main Comparison Plot Depicting Teacher APT Patterns in MR and SR



Side-by-side network models (Figure 4) and connection strengths (Table 3) further elucidated these differences. The SR network was more linear, dominated by a single strong connection between Say More and Challenge ($v = .23$). In contrast, the MR network was denser and more interconnected, forming a core triad among Press for Reasoning, Say More, and Add On, with strong links between Press for Reasoning and Say More ($v = .13$) and between Press for Reasoning and Add On ($v = .08$).

Figure 4

Network Models of APT Move Co-Occurrences in MR and SR

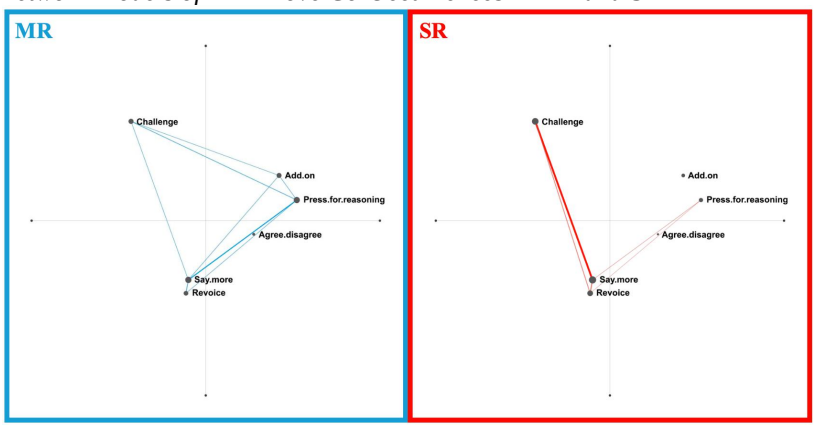


Table 3

Relative Frequencies of Co-Occurrences between Coded Dialogue Elements within a Stanza from Two Groups

MR/SR	1.	2.	3.	4.	5.	6.
1. Say More	-					
2. Revoice	0.10/0.13	-				
3. Press for Reasoning	0.13/0.07	0.08/0.06	-			
4. Challenge	0.08/0.23	0.02/0.11	0.09/0.04	-		
5. Agree & Disagree	0.01/0	0.01/0	0.01/0	0.02/0	-	

 © ISLS

6. Add On	0.08/0.04	0.03/0.03	0.08/0.02	0.07/0.05	0.01/0	-
N	15/16	6/7	10/5	6/10	2/0	6/3

In summary, the teacher in the MR condition involved shorter, more frequent student contributions and the interconnected pattern linking reasoning, elaboration, and collaboration. In the SR condition, the teacher's practice was characterized by challenging students for longer elaborations, with a network dominated by a single strong connection.

Illustrative dialogue excerpts

The collaborative and integrated nature of MR dialogic patterns

As illustrated in a representative episode from the MR condition (Table 4), the teacher integrated APT moves to facilitate a collaborative and reasoning-oriented discourse. This episode, centered on interpreting experimental data and graphs for Glauber's salt and paraffin wax, showcases how the teacher wove together multiple strategies to deepen student understanding.

Table 4

A Transcript of Classroom Dialogue with Coded APT Moves from the MR Group

No	Speaker	Utterance	APT[a]
a1	Teacher	Ethan, look at the temperature data corresponding to the melting process. What are its characteristics? Compare them and describe it in your own words.	
a2	Ethan	When heating Glauber's salt, from time 4 to…	
a3	Teacher	When heating Glauber's salt, the temperature did not change from the 4th to the 9th minute, whereas for paraffin wax, its temperature kept increasing, right? This is the difference between them. Are there any similarities?	4
a4	Ethan	Both require heating.	
a5	Teacher	Both were continuously being heated the whole time, correct? Okay, you may sit. Lucia, do you have anything to add? We have the data table now, right? So, what should we do next?	6
a6	Lucia	Plot the graph.	
a7	Teacher	The graph. Okay, please come up to the front. On the right-side blackboard, based on this table, draw what you believe is the graph of Glauber's salt temperature versus time. Also, please plot the temperature change of paraffin wax on your scratch paper. Henry, please look at the table on the blackboard and check Lucia's graph. Are there any issues?	5
a8	Henry	Her graph is correct; there are no problems.	
a9	Teacher	Do you have anything to add?	6
a10	Henry	Nothing needs to be added.	
a11	Leo	Teacher, Lucia's graph should initially start labeling from 30 degrees Celsius; she didn't mark that.	
a12	Teacher	You're saying she should have marked further up, right? Does anyone else have anything to say? Nora, do you have anything to add?	5, 6
a13	Nora	No.	
a14	Teacher	Nothing? Okay, you may sit. Lucia, please, based on the graph you drew and the table content, explain it. First, explain your graph of Glauber's salt temperature versus time relationship. Why did you draw it that way?	3
a15	Lucia	From time 0 to 3 seconds, no, 0 to 3 minutes, the temperature of Glauber's salt rose from 30 degrees to 45 degrees, and then…	
a16	Teacher	45 what?	2
a17	Lucia	45 degrees Celsius. Then, from the 4th minute until the 9th minute, the temperature of Glauber's salt remained constant at 48 degrees Celsius, so that section is a straight line. Then, after 50 degrees Celsius, or after 10 minutes, the temperature of Glauber's salt...	
a18	Teacher	Continued to increase, right?	
a19	Lucia	Yes.	
a20	Teacher	What about paraffin wax?	4
a21	Lucia	Paraffin wax is…	
a22	Teacher	The temperature was continuous?	1
a23	Lucia	Changing regularly.	

Note. [a]APT moves: 1-Say More; 2-Revoice; 3-Press for Reasoning; 4-Challenge; 5-Agree & Disagree; 6-Add On

The episode begins with the teacher using questioning to scaffold understanding, immediately challenging students to identify similarities (a3), and elevating the dialogue from description to conceptual

 © ISLS

synthesis. A defining feature emerges in the active inclusion of multiple student voices. After an initial student response (a4), the teacher explicitly invites another student to contribute (a5, Add On), framing the activity as collective inquiry. This collaborative approach is reinforced when the teacher assigns one student to draw a graph and another to review it (a7, Agree & Disagree), integrating peer evaluation into the task. Notably, a student later intervenes unprompted to correct the graph (a11), and the teacher skillfully uses this moment to solicit further input from the class (a12, Add On), reinforcing co-construction as a shared norm. The discourse then shifts back to reasoning as the teacher asks the student to justify her graph (a14, Press for Reasoning). During her explanation, the teacher scaffolds precision through Revoice (a16) and Say More (a22), while a final Challenge (a20) maintains focus on comparative analysis.

In summary, this episode exemplifies the "integrated collaborative" pattern found in MR by ENA analysis, where the teacher strategically interweaves reasoning, challenge, and collaborative moves to position students as active, co-constructing participants in scientific discourse.

Guided and linear nature of SR dialogic patterns

Table 5 (for more transcripts, see Appendix I) encapsulates the teacher's tightly guided dialogic approach in the SR condition. The episode centers on a critical juncture where a student, Stella, provides an incorrect answer, confusing the two substances (b17). The teacher's response is immediate and direct, repeating the error with a questioning tone before swiftly nominating another student, Melody, to provide the correct account (b18, Add On). This intervention prioritizes factual accuracy and maintains the lesson's narrative flow. After Melody's complete and correct response (b19), the teacher cycles back to Stella, asking if she has anything to add (b20, Add On). While this offers a chance for redemption, it functions primarily as a check for individual recall rather than an invitation for collaborative idea-building. The teacher's subsequent Revoice (b22) then scaffolds Stella's fragmented response, guiding her toward the precise scientific term.

Table 5

A Transcript of Classroom Dialogue with Coded APT Moves from the SR Group

No	Speaker	Utterance	APT[a]
b14	Teacher	Everyone, please observe the animation carefully; I will ask questions later. Watch closely. See it clearly? Let's watch it again. Initially, the alcohol burner heats, and the temperature rises. Pay attention. Stella, keep watching the animation. Think about it: in this experiment, am I heating paraffin wax or Glauber's salt?	
b15	Stella	Glauber's salt.	
b16	Teacher	Please give a retrospective account of this experimental process. As Glauber's salt is heated, what are the characteristics of its change in state, and what are the characteristics of the temperature change? Do you understand what I'm asking?	
b17	Stella	Glauber's salt turns into paraffin wax.	
b18	Teacher	Glauber's salt turns into paraffin wax?! Melody, please continue for Stella.	6
b19	Melody	When the alcohol burner starts heating, the temperature continuously rises. The Glauber's salt remains solid. When heated to a specific temperature, precisely 48 degrees Celsius, and under continued constant heating, the Glauber's salt becomes a solid-liquid coexistence state. The temperature then remains constant. After the Glauber's salt completely melts, continued heating causes the temperature to rise again.	
b20	Teacher	Okay, you may sit. Stella, do you have anything to add?	6
b21	Stella	Hmm, I also remember it was 48, and then…	
b22	Teacher	You also remember at what temperature, 48?	2
b23	Stella	At 48 degrees Celsius, the Glauber's salt begins gradually changing into this…	

Note. [a]APT moves: 2-Revoice; 6-Add On

This interaction exemplifies the directive-guided discourse pattern of the SR condition. The teacher's linear application of Add On and Revoice served to correct a specific error and steer the student back on track, prioritizing factual accuracy over collaborative reasoning. As further illustrated in Appendix I, this approach aligns with the quantified network model. The teacher's talk consistently leveraged Challenge and Say More in a sequential manner to maintain tight procedural control, effectively scaffolding understanding within the limited conceptual space defined by the single-representation animation.

Discussion

This study investigated how simulation design (MR vs. SR) shapes classroom discourse and teachers' use of APT. Framed by social cognitive theory (Bandura, 1986), which emphasizes reciprocal interactions among environment, behavior, and personal factors, we conceptualized simulation type as a key environmental variable.

Findings suggest that representational design was associated with differences in both teacher facilitation and student participation in this comparative case, illuminating how representational features of digital tools may be associated with dialogic teaching.

Key findings and contribution

ENA uncovered distinct teacher strategic repertoires across conditions. In the MR setting, APT moves formed a dense, "integrated collaborative" network centered on Press for Reasoning, Say More, and Add On, reflecting a pedagogy that connects reasoning with collaborative knowledge building and aligns with the socio-constructivist principles of APT (Michaels & O'Connor, 2015). This pattern is consistent with prior work arguing that MR environments can offer diverse semiotic resources that may support reasoning and help manage cognitive demands, potentially enabling teachers to focus on collaborative discourse (Almasri, 2022; Schwamborn et al., 2011; Trypke et al., 2024). In contrast, the SR condition elicited a sparse, linear network dominated by Say More and Challenge as a "guided challenge" pattern aimed at correcting and refining student contributions. The more limited representational affordances in the SR simulations may be linked to a narrower range of discursive moves and a more directive facilitation style in this setting. These findings extend prior work on teacher facilitation and student engagement (Bouton et al., 2025; Chen, Zhang, et al., 2020; Wang, Yang, et al., 2025) by highlighting the role of instructional tools, not just teacher or student factors, in shaping dialogic patterns. Methodologically, ENA captured structural nuances in teacher talk that frequency-based analyses would overlook, offering a clearer view of how tool affordances may relate to teachers' strategic adaptations.

At the student level, MR related to more distributed participation, with more students contributing shorter, focused responses. SR, by contrast, featured more choral responses and longer explanations from fewer students, reflecting a less student-centric dynamic. One possible interpretation is that MR may lower perceived participation barriers.

For practice, these findings imply that simulation design shapes conversational possibilities. Teachers and curriculum developers may consider selecting multi-representational tools to foster collaborative reasoning. Professional development could likewise highlight how digital environments influence discursive strategy, empowering teachers to leverage MR tools for more equitable dialogue.

Limitations and future research

This study has several limitations. First, the involvement of a single, motivated teacher ensured internal consistency for the comparative case design but limits the generalizability of the findings. Accordingly, the present findings should be interpreted as a comparative description of discursive strategy patterns across two tool environments in an information-rich case, rather than as an estimate of the prevalence of APT moves in everyday classrooms. Future research should involve multiple teachers to account for variations in pedagogical style and to explore how teacher beliefs interact with tool affordances. Additionally, while ENA powerfully models the structure of teacher talk, current tools do not fully support the statistical modeling of nested data structures (e.g., turns nested within students, within classes). Future methodological developments in this area would strengthen such analyses. Finally, the study focused on teacher strategies and overall discourse structure; future work should directly measure the corresponding cognitive and affective outcomes for students to more definitively link these dialogic patterns to learning gains. Future research could also integrate discourse analysis with student interviews to explore how students perceive and experience these different participatory dynamics

References

Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction*, *16*(3), 183–198.

Almasri, F. (2022). Simulations to teach science subjects: Connections among students' engagement, self-confidence, satisfaction, and learning styles. *Education and Information Technologies*, 27(5), 7161–7181.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice-Hall, Inc.

Bouton, E., & Asterhan, C. S. C. (2023). In pursuit of a more unified method to measuring classroom dialogue: The dialogue elements to compound constructs approach. *Learning, Culture and Social Interaction*, *40*, 100717.

Bouton, E., Yosef, D., & Asterhan, C. S. C. (2025). Differences between low and high achievers in whole-classroom dialogue participation quality. *Learning and Instruction*, *96*, 102088.

Bowman, D., Swiecki, Z., Cai, Z., Wang, Y., Eagan, B., Linderoth, J., & Shaffer, D. W. (2021). The mathematical foundations of epistemic network analysis. In A. R. Ruis, & S. B. Lee (Eds.), *Advances in quantitative ethnography* (pp. 91–105). Springer.

Brown, J. S., Collins, A. M., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, *18*(1), 32–42.

Chen, G. (2020). A visual learning analytics (VLA) approach to video-based teacher professional development: Impact on teachers' beliefs, self-efficacy, and classroom talk practice. *Computers & Education*, *144*, 103670.

Chen, G., Chan, C. K. K., Chan, K. K. H., Clarke, S. N., & Resnick, L. B. (2020). Efficacy of video-based teacher professional development for increasing classroom discourse and student learning. *The Journal of the Learning Sciences*, *29*(4–5), 642–680.

Chen, G., Zhang, J., Chan, C. K. K., Michaels, S., Resnick, L. B., & Huang, X. (2020). The link between student‐perceived teacher talk and student enjoyment, anxiety and discursive engagement in the classroom. *British Educational Research Journal*, 46(3), 631–652.

Jaakkola, T., & Veermans, K. (2020). Learning electric circuit principles in a simulation environment with a single representation versus "concreteness fading" through multiple representations. *Computers & Education*, *148*, 103811.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 159–174.

Li, Y. H., Su, C. Y., & Ouyang, F. (2024). Integrating self-explanation into simulation-based physics learning for 7th graders. *Journal of Science Education and Technology*, *33*(3), 286–299.

Magner, U. I. E., Glogger, I., & Renkl, A. (2016). Which features make illustrations in multimedia learning interesting? *Educational Psychology*, *36*(9), 1596–1613.

Marquart, C. L., Hinojosa, C., Swiecki, Z., Eagan, B., & Shaffer, D. W. (2018). Epistemic network analysis [Computer software] Version 1.7.0. http://app.epistemicnetwork.org.

Michaels, S., & O'Connor, C. (2015). Conceptualizing talk moves as tools: Professional development approaches for academically productive discussion. *Socializing intelligence through talk and dialogue*, 347-362.

Michaels, S., O'Connor, M. C., & Resnick, L. B. (2008). Deliberative discourse idealized and realized: Accountable talk in the classroom and in civic life. *Studies in the Philosophy of Education*, *27*(4), 283–297.

Pehmer, A. K., Gröschner, A., & Seidel, T. (2015). How teacher professional development regarding classroom dialogue affects students' higher-order learning. *Teaching and Teacher Education*, *47*, 108–119.

Resnick, L. B., Michaels, S., & O'Connor, M. C. (2010). How (well structured) talk builds the mind. In D. Preiss & R. Sternberg (Eds.), *Innovations in educational psychology* (pp. 163–194). Springer.

Rexigel, E., Kuhn, J., Becker, S., & Malone, S. (2024). The more the better? A systematic review and meta-analysis of the benefits of more than two external representations in STEM education. *Educational Psychology Review*, *36*(4), 124.

Schwamborn, A., Thillmann, H., Opfermann, M., & Leutner, D. (2011). Cognitive load and instructionally supported learning with provided and learner-generated visualizations. *Computers in Human Behavior*, *27*(1), 89–93.

Sedova, K., & Sedlacek, M. (2023). How vocal and silent forms of participation in combination relate to student achievement. *Instructional Science*, *51*(2), 343–361.

Tao, Y., & Chen, G. (2024). The relationship between teacher talk and students' academic achievement: A meta-analysis. *Educational Research Review*, 45, 100638.

Testa, I., & Catena, D. (2022). High school students' performances in transitions between different representations of linear relationships in mathematics and physics. *Education Sciences*, *12*(11), 776.

Trypke, M., Stebner, F., & Wirth, J. (2024). The more, the better? Exploring the effects of modal and codal redundancy on learning and cognitive load: An experimental study. *Education Sciences*, *14*(8), 872.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. *Cambridge*, MA: Harvard University Press.

Wang, P., Chen, F., Wang, D., & Chen, G. (2025). Enhancing students' dialogic reflection through classroom discourse visualisation. *International Journal of Computer-Supported Collaborative Learning*, *20*(3), 293-315.

Wang, P., Yang, C., Tong, Y., & Chen, G. (2025). Discursive engagement for intersubjectivity: Investigating the verbal–visual relationship of productive classroom talk in synchronous online lessons. *Learning, Culture and Social Interaction*, *51*, 100892.

Acknowledgments

This work was supported by Hong Kong Research Grants Council, University Grants Committee (Grant No. 17619325).

Appendix I. Transcript of Classroom from the SR Group

No	Speaker	Utterance	APT[a]
b1	Teacher	The three states of matter are interconvertible. Let's take the most common example: water. What state is water itself in?	
b2	Ss	Liquid.	
b3	Teacher	Liquid. And if it's ice?	4
b4	Ss	Solid.	
b5	Teacher	If it becomes water vapor?	4
b6	Ss	Gaseous.	
		…	
b7	Teacher	First, let's talk about how to use these experimental setups. Look at this apparatus; it's already assembled, right? But when we actually conduct the experiment, the equipment we receive definitely won't be pre-assembled. Think about it: should the order for setting up the experimental apparatus be from top to bottom, or from bottom to top? Aria. What do you think? From bottom up or top down?	
b8	Aria	From bottom up.	
b9	Teacher	Why? Why assemble from bottom up? The textbook doesn't say, you need to think for yourself. When a classmate is answering, others, please don't interrupt. Give the speaking student some time to think. If you already know the answer, keep it to yourself for now. If the other person can't answer and you want to respond, then you'll get your turn. Think about it. What is the biggest difference between assembling from top down versus bottom up?	3
b10	Aria	When assembling from the bottom up, you place the candle first, then continue installing upwards step by step. After that, you place the wire gauze and the water at the top.	
b11	Teacher	What are the advantages of doing it this way? Think about the benefits.	1
b12	Aria	Assembling this way allows for effective control of the installation space.	
b13	Teacher	Correct. Okay, you may sit. The key point is controlling the spacing. If you assemble from top to bottom, it's difficult to gauge how much space to leave, right? Introduce the functions of the other instruments. What is the purpose of the wire gauze?	

Note. [a]APT moves: 1-Say More; 3-Press for Reasoning; 4-Challenge

 © ISLS

International Society of the Learning Sciences

Supporting Students' Shared Epistemic Agency Discourse with GAI via Learning Analytics Augmented Meta-discourse

Yawen Yu, Dongguan University of Technology, yawenyu@connect.hku.hk
Can Sun, Dongguan University of Technology, sunc@dgut.edu.cn
Gaowei Chen, The University of Hong Kong, gwchen@hku.hk

Abstract: This study investigates the effectiveness of a learning analytics-augmented meta- discourse environment in fostering students shared epistemic agency discourse in a Generative AI (GAI) learning context. Two types of learning analytical tools and meta-discourse were utilized to support students' epistemic understanding of critical discourse, thus, to support students' epistemic agency discourse. A total of 48 project management students participated in the 18-week course, engaging in four online discussions and two meta-discourse sessions. Findings show that the intervention significantly improved students' epistemic understanding of critical AI discourse. Furthermore, students demonstrated progress in their shared epistemic agency discourse, moving from information sharing to more projective and regulative activities across the three phases of discussion. Students with high epistemic understanding showed a stronger tendency to build on, evaluate AI information, and create joint action plans, in contrast to low-understanding students who primarily used AI information for idea negotiation. These results underscore the critical role of pedagogical design in utilizing GAI as an epistemic tool to support collaborative knowledge advancement.

Introduction

With the advent of Generative AI (GAI) tool, researchers have been investigating the effectiveness of GAI in supporting students' higher-order thinking skills compared with traditional non-AI learning environment. Existing research has shown that GAI enhances students' creativity and critical thinking abilities (Liu et al., 2024), facilitates reflective thinking and the acquisition of conceptual knowledge (Yang et al., 2025), and promotes knowledge comprehension while reducing free-riding behavior in collaborative settings (Gong et al., 2024). Among the various skills that studies aim to enhance, few have investigated whether GAI supports students shared epistemic agency, and even fewer have explored strategies for promoting it. Shared epistemic agency refers to the dynamic process through which students collaboratively and intentionally engage with individual ideas and external information sources, constructing collective shared understanding that actively advances knowledge rather than passively accepting given information (Damsa et al., 2010). Supporting shared epistemic agency becomes important especially in the age of GAI, where information is easily accessible. Without epistemic agency, students would treat GAI-generated content as an authoritative, unquestionable entity and consume information without thoughtful consideration. Therefore, a significant challenge lies in designing learning environments that effectively utilize GAI as an epistemic tool supporting students shared epistemic agency.

While research on the use of Generative AI (GAI) to enhance shared epistemic agency is limited, existing studies have delved into this concept within traditional knowledge-building environments. Researchers utilize design components such as learning analytics and meta-discourse to support students ' epistemic agency (Tong & Chan, 2023; Yang, 2019; Yu et al., 2024; Sun et al., 2025). Tong and Chan (2023) utilized a learning analytics-augmented meta-discourse design to promote students' epistemic discourse understanding. The study, conducted within a knowledge-building context, showed that the experimental group developed a more sophisticated epistemic discourse understanding and greater domain knowledge than a comparison class. The authors argued that this meta-level understanding is a critical foundation for engaging in productive inquiry. Yang (2019) designed a reflective assessment framework to foster shared epistemic agency among low-achieving students in a knowledge-building environment. This research demonstrated that reflective assessment helped students set goals, monitor their progress, and regulate their inquiry, thereby promoting high-level epistemic agency. Yu et al.(2024) examined the effectiveness of learning analytics-augmented meta-discourse in supporting shared epistemic agency. The findings showed that the approach was effective in helping college students develop shared epistemic agency discourse, characterized by their engagement in a comprehensive discourse trajectory that included projective planning, information sharing, and creating shared understanding.

The dynamics of using learning analytics to augment teacher and students meta-discourse sessions could be concluded from the following perspectives. Learning analytical tools, such as Knowledge Building Discourse Explorer (KBDeX), function as intellectual artifacts, synthesizing students fragmentary online conversations into coherent representations and made students' invisible interaction levels and discourse structures visible (Chen & Zhang, 2016). The tangible representation of students ' progressive inquiry

 © ISLS

becomes the collective object of inquiry for subsequent reflection in a meta-discourse session, which is defined as discourse about discourse (Tong & Chan, 2023; Lei & Chan, 2018). During these sessions, students and teachers collectively review and analyze representations that demonstrate students' different levels of group interaction, participation, and discourse structures. This process of reflecting on the goals, characteristics, and processes of discourse ultimately allows the community to co-construct epistemic understanding of discourse, establishing the communal standard for productive inquiry and advancing knowledge (Tong & Chan, 2023; Tong et al., 2023). Epistemic understanding of discourse refers to a student's meta-level comprehension of the goals, standards, and strategies that support their engagement in productive discussions (Tong & Chan, 2023). According to Chinn's AIR model, this conceptualization encompasses a student's perception of three main constituent components: Aims and Goals, Standards and Ideals, and Reliable Processes and Strategies. The first component, Aims and Goals, focuses on the fundamental purpose of the discussion, which might include information sharing, collaborative problem-solving, or the collective advancement of knowledge. Building on this foundation, Standards and Ideals outline the criteria for a successful discussion, such as the maintenance of sustained inquiry, the generation of new ideas, or a focus on productive dialogue as opposed to a simple lack of arguments. Finally, Reliable Processes and Strategies detail the specific methods used to achieve the established aims and adhere to the ideals, involving practices like building on each other's contributions, synthesizing diverse ideas, and identifying knowledge gaps within the group's current understanding.

The present study

This study identifies two key research gaps. First, among the studies, there is a lack of research that investigates the effectiveness of a learning environment that integrates learning analytics and meta-discourse to foster students shared epistemic agency in a GAI learning context. Second, the underlying mechanism of using learning analytics- augmented meta-discourse to support students' shared epistemic agency in GAI learning environments is yet to be explored. Tong & Chan (2023) identified that learning analytics augmented meta-discourse session supports low achieving students' development of epistemic understanding of discourse and the positive correlation between epistemic understanding and productive knowledge building discourse. However, it remains unclear whether these findings hold true within the GAI learning environment.

To address these research gaps, we adopted the existing pedagogical design approach of using learning analytics augmented meta-discourse to support students shared epistemic agency. Rather than directly targeting the concept of shared epistemic agency, we opted to support students' epistemic understanding of critical discourse for following reasons. First, the concept of shared epistemic agency is rarely studied in China which made students feel unfamiliar with this notion. Second, the concept of critical discourse which is conceptualized by Newman et al (2002) involves similar collaborative process as shared epistemic agency (Damsa et al., 2010). Both frameworks highlight the importance of social and collaborative actions. Shared epistemic agency is defined as "collaboratively and intentionally" engaged process of constructing collective understanding, which aligns with the view in Newman et al. (2002) that critical thinking process is a dynamic activity developing through social interaction where meaning is socially constructed.Thus, promoting students' epistemic understanding of critical discourse serves as an effective way to support their engagement in the processes of shared epistemic agency. The study systematically supported students' different dimensions of epistemic understanding with different design sessions. The first session fostered students' meta-level comprehension of the Aims and Goals and Standards and Ideals of productive discourse, focusing on the importance of sustained engagement with AI-generated information. The second session utilized the same design features to develop students' knowledge of Reliable Processes and Strategies, enabling them to grasp the specific techniques required for engaging in critical discourse.

In this study, we investigate the effectiveness of using learning analytics-augmented meta-discourse to support students epistemic understanding of critical discourse, thus to support their engagement in shared epistemic agency discourse. We ask following research questions. 1.To what extent does the learning analytics- augmented meta-discourse support students' development of shared epistemic agency discourse? 2.To what extent does the learning analytics-augmented meta-discourse support students' epistemic understanding of discourse?3.How students' with different epistemic understanding demonstrate different levels of shared epistemic agency discourse features?

Method

Participants and instructional contexts

This study involved 48 students, including 40 males and 8 females. The course, titled "Evaluation of Engineering Investment and Financing", was attended by students majoring in project management. All participants were in their second year of study and had successfully completed the prerequisite courses prior

to the class. The instructor had 15 years' teaching experience and knew the students well. The semester spanned 18 weeks and was structured into three sections, each lasting six weeks and concentrating on different topics. The first section (weeks 1-6) introduced fundamental concepts of project investment and financing strategies. The second section (weeks 7-12) concentrated on project investment, while the third section (weeks 13-18) focused on financing. During the first and third sections, the instructor employed lecture-based instruction, whereas the second section emphasized inquiry-based learning. Throughout the second section, students acquired foundational knowledge in writing investment plans during class and collaborated in small groups to draft their investment plans outside of class. Students voluntarily formed groups, with each group consisting of six members, resulting in a total of eight groups.

From the eighth week, each group presented their milestone achievements in drafting their investment plans during class. Before these presentations, students participated in online asynchronous discussions via WeChat, where they shared information and deliberated on the analyses necessary to complete their investment plans. Each group established a WeChat group visible only to its members, resulting in a total of eight groups created. Students were encouraged to engage with GAI to ask questions, summarize key concepts, and identify issues or errors in their calculations. Furthermore, students were required to document their interactions with AI throughout these discussions. In total, four online discussions were conducted, each addressing essential elements of an investment plan. The first discussion, held before the 8th week, focused on the project overview and market analysis. The second discussion took place before week 10 and dealt with project planning and some financial analysis. The third discussion, occurring before week 11, concentrated on both financial analysis and risk assessment, while the final discussion explored the topic of proposing countermeasures. There was a break during the 9th week for a fall recess, and interventions were implemented during the 8th and 11th weeks of class. The instructor hosted two learning analytics-augmented meta-discourse sessions to enhance students' shared epistemic agency. The two sessions divided the online discussions into three phases: the first phase encompassed the initial discussion, the second phase included both the second and third discussions, and the third phase comprised the final discussion.

Pedagogical design

Meta-discourse session I

The primary goal of the first design session is to help students develop an epistemic understanding of the importance of critically engaging with AI. To effectively support this, we implemented following design components. First, the teacher selected online discussions from two contrasting groups: one group where students actively built upon each other's ideas, including AI information and another where responses predominantly came from students' agreements with AI information. Second, learning analytics visualizations were utilized to illustrate the levels of engagement in critical AI discourse. Two networks were generated from the raw discussion data: one displayed a centralized structure, indicating that most students were responding exclusively to the AI information, while the other showcased a decentralized structure, with multiple students actively engaging and developing each other's ideas (see Figure 1.a; a clearer version of the figure could be obtained by visiting following link: https://drive.google.com/file/d/1QtRt9z1se-fuiRQsHA5cfAvzr8ubQP5F/view?usp=sharing). Third, the teacher facilitated meta-discourse session, encouraging students to identify the contrasting features of the two networks, including the shape of the network and the thickness of the connections between students. Additionally, the teacher articulated what these patterns reveal about students' online discourse. To enhance comprehension, snippets of the discussions were presented, and network features were mapped to specific discourse segments (see Figure 1.b; a clearer version of the figure could be obtained by visiting following link: https://drive.google.com/file/d/1QtRt9z1se-fuiRQsHA5cfAvzr8ubQP5F/view?usp=sharing).

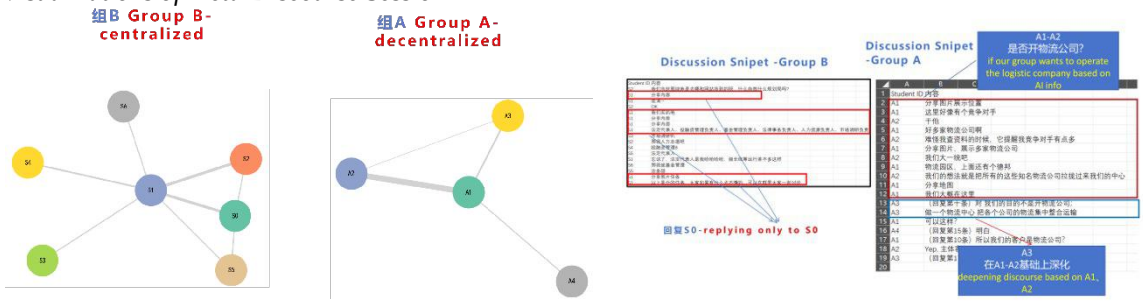
Meta-discourse design session II

In the second design session, students' epistemic understanding is supported from processes and strategies of critical discourse engagement. One major aim is to support students' understanding of what types of strategies they could engage in conducting critical discourse. We employed similar design components as the first design session, namely, contrastive group discussions, learning analytics and meta-discourse. In this design session, the teacher first selected two contrastive group discussions which are group A and B (see Figure 2a-b; a clearer version of the figure could be obtained by visiting following link: https://drive.google.com/file/d/1QtRt9z1se-fuiRQsHA5cfAvzr8ubQP5F/view?usp=sharing). Students from the two groups showed different levels of critical use of AI information. In this session, ENA was employed to create learning analytical visualizations to support students' reflective practices in following meta-discourse sessions (see Figure 2c; a clearer version of the figure could be obtained by visiting following link: https://drive.google.com/file/d/1QtRt9z1se-fuiRQsHA5cfAvzr8ubQP5F/view?usp=sharing). To create ENA figures, the teacher made similar attempts as data analysis, including following steps. First, the teacher coded students' online data with the coding scheme from Newman et al (1995) which aimed to measure how college

students demonstrated critical thinking in computer-supported group learning with adaptations (see table 1). The teacher coded segments of students' discussions from both groups. In total, the teacher coded 156 notes, in which 76 were from group A and 80 were from group B. The teacher then explained what different components in one ENA graph indicate. For example, what do the dotted squares, the squares, the codes and the edges that connected the codes indicate. The teacher then guided students to notice different locations of the centroids and explained to the students the meanings of different centroids locations. The teacher firstly explained that the centroids of the two groups located in different directions of the X-axis showed students' different discourse patterns. Furthermore, the teacher explained that in group A (shown as red lines and squares in figure 3c), students engaged in discourse types as evaluating solutions, providing information related to the question, and sustaining the inquiry by asking more questions while in group B (shown as blue lines and squares in figure 3c), students engaged in extending AI information, providing initial solutions to the problem and planning for the inquiry process. The teacher also showed snippets of students' online discussions and demonstrated the flow of discussions. For example, the teacher showed that students firstly referred to AI information, then evaluated AI information. Students then evaluated how they should change their information searching strategies based on the information they have searched. Similarly, the teacher also referred to the discussion snippets from group B and explained that the discussion dynamics showed that students appeared to neglect incorporating information from AI nor asking further questions from AI to improve their current answers (see Figure 2). From this, students developed their epistemic understanding of critical AI discourse further, identifying that critical AI discourse involves evaluating, identifying information obtained from AI resources, monitoring object advancement, adjusting group strategies and checking validity.

Figure 1

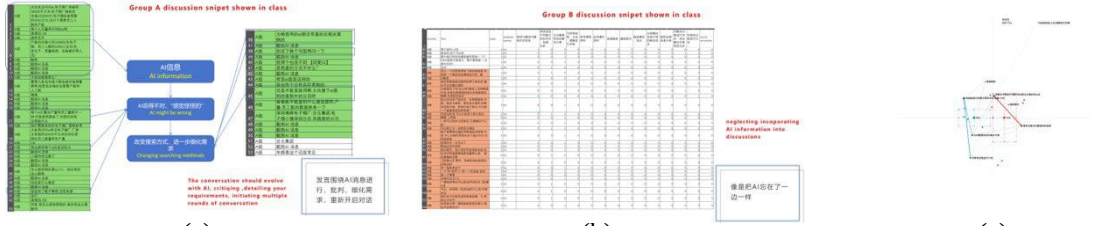
Visualizations of Meta-Discourse Session I



(a) (b)

Figure 2

Visualizations of meta-discourse session II



(a) (b) (c)

Table 1

Students' Critical Discourse Dimensions

Dimensions	Definitions
new ideas for discussion	students brought up new ideas to discuss
Information critique	students evaluated, extended and analyzed information
new solutions to the problem	students provided new solutions to the problem

 © ISLS

connecting with personal experiences and understanding	Students came up with solutions by connecting personal experiences and understanding
referring to course material	Students referred to course material to identify solutions to the problem
use relevant outside material	Students used outside course material to solve the problem at hand
clarifying confusion	Students asked questions to clarify confusions
sustaining the inquiry	Students asked new questions to sustain the inquiry
connecting facts and ideas	Students connected facts and ideas to rise above new ideas
producing new ideas from collected information	Students concluded new ideas by summarizing information they collected.
providing evidence and example	Students provided evidence and examples for their proposed ideas
solution evaluation	evaluating problem solutions and listing advantages and disadvantages
planning inquiry process	Students planned the inquiry based on their current progress

Data collection and analytical methods

Data collection

The primary data sources for this study were WeChat discussion posts made by students. As previously mentioned, students engaged in four online discussions across three phases, resulting in a total of 981 notes. Specifically, they contributed 83 notes in Phase 1, 620 notes in Phase 2, and 278 notes in Phase 3. All online discussion posts were written in Chinese. This student-generated online discussion data was utilized to address the first research question.

To explore students' epistemic understanding of critical AI discourse, three questions were designed to measure students' understanding. The questions were designed based on Chinn's conceptualization of students' epistemic cognition (Chinn et al., 2014). The first question addressed the standards of effective discourse: "What do you think are the features ofa good human-AI interaction?" The second question sought to measure students' understanding of the objectives behind human-AI interaction: "What do you believe is the purpose of human-AI interaction?" Lastly, the third question focused on the strategies employed to foster productive discourse: "what types of strategies did you use to enhance your engagement with AI-related information?" These questions were used to assess students' epistemic understanding both at the beginning of the semester and at the end of the intervention. For both assessments, students submitted their responses by scanning a quick-response code provided by the teacher, resulting in a total of 46 collected responses for each assessment. Prior to data analysis, all personally identifiable information was removed. The study obtained approval from the university's Institutional Review Board.

To investigate how learning analytics-augmented sessions facilitate the development of epistemic understanding among students, following data sources were collected. We conducted interviews with one driving question, namely, to what extent do you think the strategies you learned in class support your future studies, especially your interactions with AI? In total, 46 students participated in the interview. All the interviews were recorded and transcribed for future analyses. In addition to the interview data, students' learning artifacts were collected, comprising their investment plans and daily reflections. In the investment plans, students were prompted to evaluate the degree to which their interactions with generative AI contributed to the process of crafting the plan.

Analytical methods

To address the first question, we conducted following analyses. First, we exported students' online discussion notes into a spreadsheet and then imported them into NVivo 12 for further examination. Next, we carried out content analysis, adapting Damsa's dimensions of shared epistemic agency to characterize students' discourse (2010). In addition to the original 24 actions, which were grouped into four categories to reflect students' engagement in shared epistemic agency discourse, we included new actions that indicated students' interactions with AI. These two dimensions involved sharing information from AI and criticizing AI-generated information (see Table 2). One researcher coded the data, while a second researcher performed reliability checks. We achieved a reasonable reliability score of 0.897 and discussed any inconsistencies found in our coding results. Following this, we conducted an epistemic network analysis (ENA) based on the methods described by Shaffer et al. (2016). In this study, we sought to analyze the extent to which students developed shared epistemic agency discourse over the three phases. Therefore, we used whole conversations as the stanza size, with "phases" serving as the activity unit.

Table 2
Shared Epistemic Agency Discourse Dimensions

Categories of Actions	Actions	Abbreviation in ENA
Creating Awareness	Identifying problems	ca-ip
Creating shared understanding	Sharing information from AI	csu-siAI
	Sharing information from other resources	csu-si
	Sharing ideas and knowledge	csu-sik
	Criticizing AI information	csu-cai
	(re)framing the problem	csu-rm
Generative collaborative actions	Generating new ideas	gca-gni
	Engaging in collaborative uptake of ideas	gca-eicui
	Negotiating new ideas	gca-nni
	(Re)framing the object	gca-rto
	Engaging in shared construction of objects	gca-escoo
	Revising ideas and objects	gca-riod
Projective	Setting common goals	p-scg
	Creating a joint plan of actions	p-cajpa
Regulative	Monitoring object advancement	r-moa
	Coordinating object-related activities	r-cora
	Reflecting on actions, ideas and problems	r-roaip

To address the second question regarding how students with varying levels of epistemic understanding exhibited different levels of shared epistemic agency discourse, we adopted following approaches. Firstly, we developed coding schemes (see Table 3 for details). The coding scheme was developed using a theory- and data- driven approach adapted from the AIR model, critical thinking measurement scales developed by Newman et al and the shared epistemic agency discourse dimensions developed by Damsa et al (2010). Second, we coded the data. One researcher coded the data, and a second researcher conducted reliability checks. We achieved reasonable reliability (0.856) and discussed the inconsistencies in our coding results. Consequently, each student received three scores, ranging from 1 to 3, and we calculated a total score for each student. Both a pre-test and a post-test total score were assigned for each individual. Finally, a pairwise t-test was carried out to determine whether students gained a deeper epistemic understanding of critical AI discourse throughout the intervention.

Table 3

Coding dimensions of students ' epistemic understanding

Questions	Level 1	Level 2	Level 3
What do you think are the features of a good human-AI interaction?	Students vaguely mentioned the importance of using AI information to provoke further thoughts	Students identified the importance of asking clear and detailed questions	Students identified the importance of iteratively asking AI questions
What do you believe is the purpose of human-AI interaction?	Students articulated the purpose of human-AI interaction is to seek answers	Students articulated the purpose of human-AI interaction is to diversify ideas	Students articulated the purpose of human-AI interaction is to synthesize ideas
What types of strategies did you use to enhance your engagement with AI-related information?	Students just mentioned the importance of asking questions.	Students identified the importance of sustaining the inquiry	Students mentioned the importance of synthesizing answers and ideas generated from AI

Based on the students' total scores in their post-tests, we ranked their levels of epistemic understanding, identifying the top 13 as high-epistemic understanding students and the bottom 13 as low-epistemic understanding students. With students' different epistemic understanding levels, we investigated the third research question. We conducted a second round of ENA analysis to understand how students' epistemic understanding shaped their shared epistemic agency discourse. We used students' different epistemic understanding levels as activity units, actions of shared epistemic agency discourse as codes, and whole conversation as a stanza.

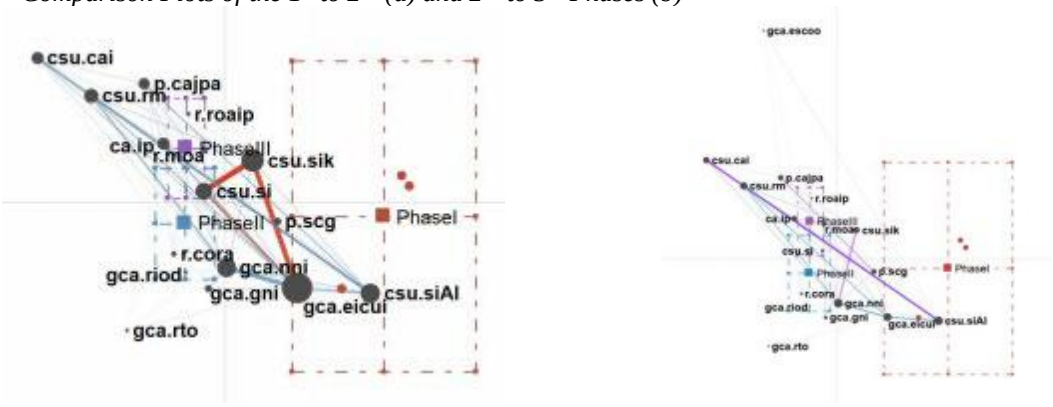
Findings

To what extent does the learning analytics-augmented meta-discourse support students' development of shared epistemic agency across the three phases?

To better understand how students' shared epistemic agency discourse progressed across the three phases, we conducted ENA analysis. The distribution of mean centroids and nodes, which represent different discourse moves, is shown in Figure 4a-b. In the first phase, the centroids ofthe mean epistemic networks were located to the right of the y-axis, indicating that students were engaging in creating shared understanding, sharing information and knowledge, and collaboratively uptake of ideas. While in the second and third phases, the centroids ofthe mean epistemic networks were located to the left of the y-axis, indicating students were engaging in more diverse types of shared epistemic agency discourse. Specifically, during the second phase, the centroids located to the bottom of the y-axis, indicating students were more engaged in generating collaborative actions but fewer regulative activities. In the third phase, students engaged in projective and regulative activities, while there was a decrease in creating shared understanding activities. Statistical analysis revealed significant differences in the mean epistemic network of students' shared epistemic agency discourse moves between the first and second phases along the x-axis (t = 8.02, p < .01), between the first and third phases (t = 8.93, p < .01), and between the second and third phases along the y-axis (t = 2.67, p < .05).

Figure 3

Comparison Plots of the 1st to 2nd (a) and 2nd to 3rd Phases (b)



(a) (b)

To what extent does the learning analytics-augmented meta-discourse support students' epistemic understanding of discourse?

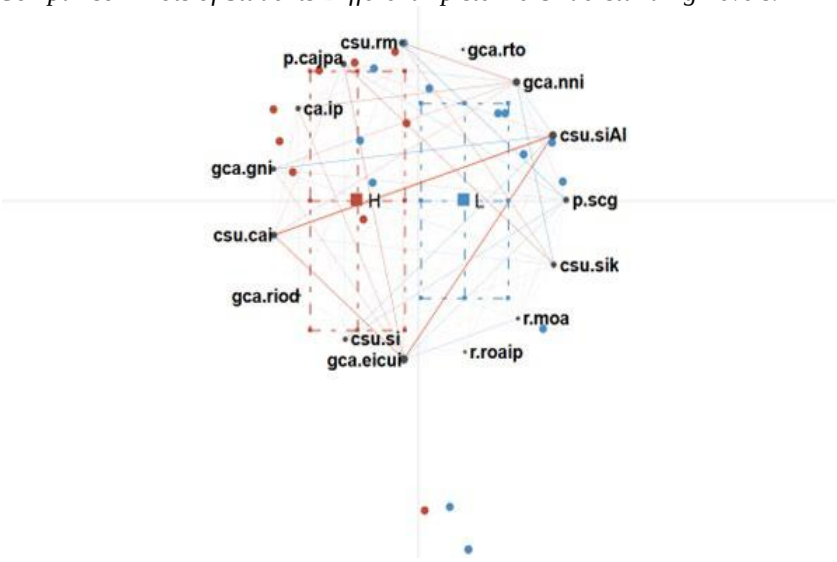
The pairwise t-test results showed that students epistemic understanding of critical AI discourse significantly improved from the pre-test to the posttest (pre-test mean = 3.27, posttest mean = 5.80, t = -11.206, p < .01).

How students with different epistemic understanding levels demonstrate different levels of shared epistemic agency discourse features?

Students were grouped as high understanding and low understanding groups based on their epistemic understanding scores. Specifically, students were regarded as high epistemic understanding students if their epistemic understanding scores were one standard deviation above the mean and as low epistemic understanding students if their score were one standard deviation below the mean. Consequently, 13 students were identified as belonging to the high epistemic understanding cohort (with mean score as 7.92), while an equal number of 13 students constituted the low epistemic understanding group (with mean score as 3.85). To better understand how students epistemic understanding of critical AI discourse shape their shared epistemic agency discourse engagement, we conducted another round of ENA analysis. We combined students' three phases of discussions and examined how students with different epistemic understanding showed different patterns of shared epistemic agency discourse. As shown in figure 5, the centroid of the high-epistemic understanding students is located to the left of the y-axis, whereas the centroid of the low epistemic understanding students is located to the right of the y-axis. The t-test score showed the difference in students' shared epistemic agency discourse (t = 3.70, p < .01). The comparison plot demonstrates that high epistemic understanding students tend to build on, evaluate AI information to inform their decision-making process while low understanding students utilized AI information for the purpose of idea negotiation without integrating this information into their decision-making processes.

 © ISLS

Figure 4.
Comparison Plots of Students Different Epistemic Understanding Levels.



Discussion

Our findings demonstrate that a Learning Analytics (LA)-augmented meta-discourse environment significantly improves students' epistemic understanding of critical AI discourse. This aligns with Tong et al. (2023) regarding the utility of meta-discourse in fostering productive inquiry. While Tong noted high epistemic understanding correlates with traditional knowledge-building (e.g., synthesizing), our study reveals its specific role in the GAI context: students with high understanding evaluate AI outputs to create joint action plans, whereas those with low understanding negotiate ideas without substantive integration. Furthermore, consistent with Yu et al. (2024), we observed a shift from information sharing toward projective and regulative activities. Notably, engaging with GAI appears to catalyze these higher-order moves more intensely than traditional environments. This work enriches the notion of shared epistemic agency for the GAI era. We establish that supporting students' epistemic understanding of AI discourse is a primary mechanism for promoting agency. Methodologically, we demonstrate that integrating tools like ENA—not just for post-hoc assessment, but to generate "objects of inquiry" for reflection—enables students to visualize and refine collective discourse patterns. Practically, we offer a clear pathway for cultivating shared agency by establishing critical AI discourse as a foundational collaborative practice.

References

Chen, B., & Zhang, J. (2016). Analytics for knowledge creation: Towards epistemic agency and design-mode thinking. Journal of Learning Analytics, 3(2), 139–163.

Chinn, C. A., Rinehart, R. W., & Buckland, L. A. (2014). Epistemic cognition and evaluating information: Applying the AIR model of epistemic cognition. In D. Rapp & J. Braasch (Eds.), Processing inaccurate information: Theoretical and applied perspectives from cognitive science and the educational science (pp. 425–453). MIT Press.

Damşa, C. I., Kirschner, P. A., Andriessen, J. E. B., Erkens, G., & Sins, P. H. M. (2010). Shared epistemic agency:An empirical study of an emergent construct. The Journal of the Learning Sciences, 19(2), 143–186. https://doi.org/10.1080/10508401003708381

Gong, R., Jiang, R., Guo, C., Hu, W., & Li, Y. (2024). Roles emerging during the knowledge construction process in collaborative learning: Does a generative AI-support chatbot matter? In Proceedings of the 2024 the 16th International Conference on Education Technology and Computers (ICETC 2024) (pp. 1–9). ACM. https://doi.org/10.1145/3702163.3702165

Lei, C., & Chan, C. K. (2018). Developing metadiscourse through reflective assessment in knowledge building environments. Computers & Education, 126, 153-169.

Liu, Y.-L. E., Lee, T.-P., & Huang, Y.-M. (2025). Enhancing student engagement and higher-order thinking in human-centred design projects: the impact of generative AI-enhanced collaborative whiteboards. Interactive Learning Environments. Advance online publication.https://doi.org/10.1080/10494820.2025.2496973

Newman, D. R., Webb, B., & Cochrane, C. (1995). A content analysis method to measure critical thinking in face-to-face and computer supported group learning. Interpersonal Computing and Technology: An Electronic Journal for the 21st Century, 3(2), 56–77.

Sun, C., Huang, B., Zheng, Y., Wang, Y., & Yu, Y. (2025). Supporting students' collaborative knowledge building via learning analytics supported metacognitive meeting. Interactive Learning Environments. Advance online publication.https://doi.org/10.1080/10494820.2025.2509279

Shaffer, D. W., Hatfield, D., Svarovsky, G. N., Nash, P., Nulty, A., Bagley, E., ... & Mislevy, R. (2009). Epistemic network analysis: A prototype for 21st-century assessment of learning. International Journal of Learning and Media, 1(2), 33-53.

Tong, Y., & Chan, C. K. K. (2023). Promoting knowledge building through meta-discourse and epistemic discourse understanding. International Journal of Computer-Supported Collaborative Learning, 18(3), 353– 391.https://doi.org/10.1007/s11412-023-09403-2

Tong, Y., Yang, C., & Chen, G. (2023). A visual learning analytics approach for knowledge building: Impact on students' epistemic understanding of discourse, productive inquiry and domain knowledge. British Journal of Educational Technology, 55(3), 992–1019.https://doi.org/10.1111/bjet.13409

Tong, Y., Yang, C., Wang, P., & Chen, G. (2024). Promoting low-achieving students' epistemic understanding of discourse: A video-based visual learning analytics approach. Interactive Learning Environments, 32(10), 5560–5576.https://doi.org/10.1080/10494820.2023.2296517

Yang, G., Rong, Y.-D., Wang, Y.-L., Zhang, Y.-Y., Yan, J.-J., & Tu, Y.-F. (2025). How generative artificial intelligence supported reflective strategies promote middle school students' conceptual knowledge learning: An empirical study from China. Interactive Learning Environments. Advance online publication. https://doi.org/10.1080/10494820.2025.2521339

Yang, Y. (2019). Reflective assessment for epistemic agency of academically low-achieving students. Journal of Computer Assisted Learning, 35(4), 459–475.https://doi.org/10.1111/jcal.12343

Yu, Y., Tao, Y., Chen, G., & Sun, C. (2024). Using learning analytics to enhance college students' shared epistemic agency in mobile instant messaging: A new way to support deep discussion. Journal of ComputerAssisted Learning, 40(3), 1166–1184. https://doi.org/10.1111/jcal.12941

International Society of the Learning Sciences

Collaboration and Competition in Educational Gameplay: Strategic Student Decision-Making

Tianyu Ma, University of Miami, txm623@miami.edu
Jennifer Kahn, University of Miami, jkahnthrone@miami.edu

Abstract: This study examines how middle school students engaged in strategic decision-making when collaboration and competition were tightly intertwined in an educational gameplay setting. Drawing from a design-based research study of an afterschool program, six students played Isles of Ilkmaar, a multiplayer game designed to support data science learning. Using Activity Theory as an analytic lens, we traced how students mobilized tools, interpreted rules, and negotiated roles in the game's diner competition. Analysis of three gameplay episodes showed students adapting in real time, moving from tool learning to contested claims and finally to rule negotiation. Findings highlight that collaboration and competition were not separate modes of activity but operated together, with conflicts often catalyzing adaptation and role change. This work contributes to CSCL by showing how collaborative learning can be sustained under competitive pressures, offering implications for the design of educational environments that integrate both cooperative and adversarial elements.

Introduction: Strategic action in collaborative-competitive gameplay

Real-time educational gameplay often requires students to engage in both collaboration and competition. When balanced, these dynamics can strengthen participation and learning (Qiao et al., 2024; Wang & Huang, 2023). Collaboration supports problem-solving and shared progress, while competition drives persistence and performance (Plass et al., 2013). Rather than opposites, prior research shows that collaboration and competition often intersect, shaping how students approach shared tasks (Pareto et al., 2012).

In this study, we examine collaboration and competition in an educational multiplayer game world. We draw from a design-based research study in which middle school girls in an afterschool club played Isles of Ilkmaar, a newly developed educational game to support data science learning. Isles of Ilkmaar is a multiplayer game where players generate, analyze, and share data about their own gameplay while building relationships with creature inhabitants and solving mysteries across interconnected islands. The game was developed through a co-design process involving youth, researchers, and a gaming studio to create a socially engaging data science experience (Radke et al., 2024).

In Isles of Ilkmaar, players have multiple paths to success. Collaboration is not required, but gameplay is often more efficient when players coordinate their efforts. For instance, players can trade foraged items or divide responsibilities into a diner space where they prepare food for creatures, combining strengths to speed up progress. At the same time, players can view their own and others' data in the game, enabling competition across objectives. This design meant collaboration offered advantages, while competition introduced urgency that shaped decisions.

This paper investigates how a group of middle school students (n = 6) participating in a STEM afterschool program managed strategic decision-making under collaborative and competitive pressures during a two-hour after-school gameplay session of Isles of Ilkmaar. In the study session, three groups joined an informal team-based competition at Isles of Ilkmaar's diner (Figure 1). In the diner, players prepared and served meals for creatures, requiring teammates to coordinate while simultaneously competing with rivals. Competition was guided by a leaderboard data display that showed the foods produced and their ratings from creatures. The leaderboard was not a fixed mechanic, but an improvised system created by students using the game's data displays, which turned available information into a running tally of performance.

Figure 1
In-Game Diner Spaces in Isles of Ilkmaar (a) Kitchen, Where Players Prepare Foods With Ingredients in a Shared Icebox (b) Heat Lamp Station, Where Finished Food is Held Before Serving. (c) Serving Area, Where Players Deliver Meals Based on Seating.

 © ISLS


(a)


(b)



(c)

Our research question asks: *How do middle school students engage in strategic decision-making under collaborative and competitive pressures during educational gameplay?* This study examines how students make moment-to-moment strategic decisions by coordinating tools, roles, and rules under simultaneous collaborative and competitive pressures in gameplay. To address this, we apply Activity Theory (Engeström, 1987; Karanasios et al., 2021), focusing on how students use tools, adapt roles, and respond to rules in and out of the digital game world, when cooperation and rivalry intertwine. By situating decision-making in social and technological contexts, this study shows how gameplay reveals the micro-processes that support collaborative learning.

Literature review: Collaboration and competition in educational gameplay

Collaboration in serious games supports shared problem-solving and knowledge construction through cognitive and socioemotional interactions, shared artifacts, and the interplay of technology and pedagogy (Xia et al., 2025). In game-based learning, collaboration can reduce cognitive load, enhance reasoning, and deepen understanding, but these benefits depend on peer interactions being intentionally structured around shared knowledge goals (Moon & Ke, 2020). It also requires regulation of cognitive, motivational, and emotional processes, which tools such as scripts and awareness systems can support (Miller & Hadwin, 2015). Left unstructured, collaboration often leads to superficial exchanges; however, defining roles and incorporating conflict-based tasks can foster deeper interaction and improve outcomes without reducing motivation (Van der Meij et al., 2011).

Competition in educational games can motivate but also introduce challenges. When framed constructively, it encourages persistence and can improve coordination and group performance (Chen & Hwang, 2017). At the same time, it can create uneven outcomes and generate anxiety for lower-performing students (Cagiltay et al., 2015; ter Vrugte et al., 2015). It may also strain peer relationships and limit cooperative problem-solving, making facilitation critical to prevent negative effects (Yu, 2001; Almås & Giæver, 2024).

These dynamics are often embedded together in digital game-based environments, where collaboration supports teamwork and competition creates urgency and shapes resource management (Qiao et al., 2024). Related work on gamification and collaborative learning systems shows that structured game mechanics and shared digital environments can scaffold coordination and participation across learners (Deterding et al., 2011). When both are present, student players not only coordinate tasks but also balance pressures from rival groups, producing complex patterns of participation (Moon & Ke, 2020).

Beyond structured roles, other social interactions can reshape collaborative and competitive dynamics and impact learning. Humor, peer-led negotiation, and creative tool use can transform how tasks unfold, often extending beyond curricular goals (Kim & Ho, 2018). Such practices include revisiting goals, adjusting strategies on the fly, and introducing playful exchanges such as spontaneous challenges or shared jokes that sustain engagement and foster ethical reflection (Pervez & Orchard, 2025). These adaptations highlight collaboration and competition as evolving features of gameplay rather than fixed conditions.

Taken together, prior research shows that collaboration and competition interact fluidly in digital game-based learning environments. They shape how students pursue goals and interpret rules, tools, and social dynamics. This backdrop frames our analysis of how these dynamics unfolded in our gameplay setting, where students engaged in data practices such as analyzing and making inferences from datasets and displays. Our study examines how collaboration and competition drive decision-making that supports social, personally meaningful learning with data. By situating these dynamics within data science learning, our research extends prior work by showing how they shape participation and mediate how students generate, interpret, and apply data in real time.

Theoretical framework: Activity theory

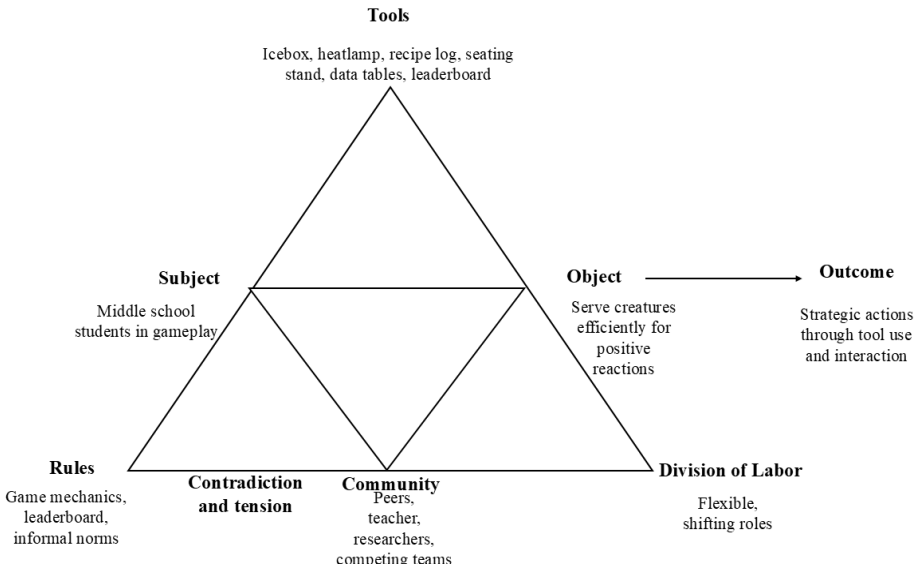
Given the role of rules, tools, and social dynamics in shaping interactions and learning experiences in multiplayer educational game environments, this study adopts Activity Theory to analyze strategic decision-making under collaborative and competitive pressures. Activity Theory, developed by Engeström (1987) and extended to digital learning contexts (Karanasios et al., 2021), views purposeful activity as the unit of analysis. In CSCL environments, digital tools and shared representations mediate collaboration, providing a lens for examining how artifacts shape interaction (Suthers, 2006). Other CSCL research has illustrated the interdependence of tools,

goals, rules, and roles in structuring human action (Fischer et al., 2013; Danish et al., 2020), making Activity Theory well-suited for analyzing student decision-making in game-based environments.

Figure 2 presents the Activity Theory framework used in this study, adapted from Engeström (1987). Applied to the Isles of Ilkmaar game world and informal club context, the subject is middle school students engaged in gameplay, with the object defined as serving creatures efficiently within the diner setting. In-game tools such as recipe log, icebox, heat lamp, data tables, and leaderboard mediate student action, while rules include both formal game mechanics and informal social norms. The community comprises peers, teachers, researchers, and rival teams, and roles are flexibly distributed. Within Activity Theory, contradictions are systemic tensions that arise within or between these elements (Engeström, 1987). Here, contradictions refer to moments when collaborative and competitive goals conflict or overlap, prompting students to reorganize strategies and roles. Such tensions are productive rather than disruptive, marking points where collective activity transforms through negotiation and adaptation.

Figure 2

Activity Theory Framework Guiding This Study



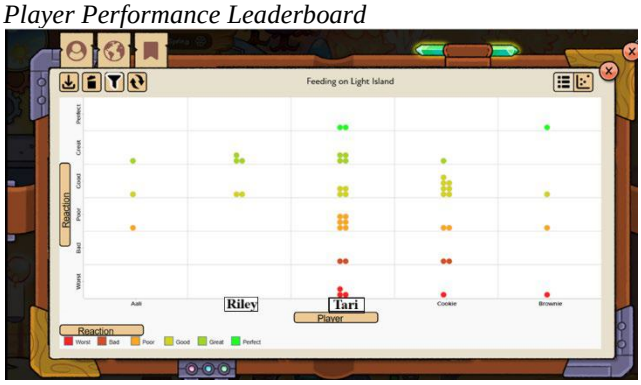
Building on this framework, we turn to the gameplay data to examine how collaborative and competitive dynamics unfolded in practice. Activity Theory frames how students used tools, interpreted rules, and negotiated roles in real time. In the following findings, we analyze selected gameplay episodes to show how talk, action, and data practices were intertwined, offering a micro-level view of strategic decision-making under the dual pressures of cooperation and rivalry.

Methodology

The larger project took place during an eight-week afterschool program with five African American girls ages 11–12 who played *Isles of Ilkmaar* together in researcher-facilitated sessions. This analysis focuses on the final two-hour session, which included solo, paired, and team play. Students accessed the game individually on laptops using their own avatars before forming pairs (one student teamed with the club teacher) for a timed diner competition. Due to one participant leaving early, our analysis focuses on one team of two students and on one student–teacher team. The one-hour rule required players to prepare and serve as many accurate dishes as possible, with winners determined by speed and creature ratings (ranging from worst to perfect, though only good, great, and perfect were scored, weighted 1–3) shown on the in-game leaderboard (Figure 3). Students coordinated tasks for accuracy and efficiency while competing for rankings. All participants had prior experience with *Isles of Ilkmaar* from earlier sessions, which involved interpreting data tables, managing ingredients and tools, and responding to creature preferences.

 © ISLS

Figure 3
Player Performance Leaderboard



Data collection for the session involved two researchers video recording each group with dedicated cameras and an audio recorder. Cameras captured on-screen activity and embodied interactions, while audio ensured dialogue could be synced with video. One researcher also had her own avatar (Author 1) and entered the game environment to follow players' avatars directly, creating a parallel record of in-game actions. This setup captured digital choices and coordinated activity.

Analysis was conducted using ATLAS.ti (Derry et al., 2010) to identify patterns in audiovisual data. A content log (Jordan & Henderson, 1995) was developed to mark moments when strategy or coordination visibly shifted. Three gameplay episodes where rivalry was most observable were selected. These episodes were transcribed into turn-based tables with speech and in-game actions recorded line by line. Following Saldaña (2016), we applied structured coding, beginning with Activity Theory categories as deductive codes and refining them inductively through iterative analysis to capture unexpected practices or interactions.

Coding followed Activity Theory, beginning deductively and then refined inductively to reflect unexpected practices. Two episodes (Episode 2 and Episode 3) were modeled as sequential decision tables and visualized as networks to trace how strategies unfolded in real time. Episode 1 is included for context but not analyzed with the same depth of modeling as Episodes 2 and 3. This synthesis made it possible to examine how students coordinated collaboratively under competitive pressure, revealing how choices were shaped by tools, rules, and community interactions within a shared digital environment.

Our network maps followed an Activity Theory coding scheme to analyze how students coordinated tasks during gameplay. The object (red) was the goal of efficiently serving creatures, often achieved through improvisation. The subject (orange) represented student decision-makers, such as Riley, who shifted between gathering and cooking based on team needs. The tools (green)—recipe log, heat lamp, icebox, seating stand, and leaderboard—mediated these actions; for instance, Brownie used the heat lamp to serve a dish before another team could, gaining a competitive edge. The rules (purple) combined system constraints, like storing ingredients before cooking, with informal norms such as fairness in sharing resources. The community (orange) included peers, teachers, researchers, and rival teams; for example, Cookie's public leaderboard comments shaped how others responded. The division of labor (dark green) was fluid, with roles shifting as needed. Contradictions and tensions (brown) marked moments when collaboration and competition collided—such as Riley hoarding ingredients to prevent theft—revealing how players negotiated fairness and control. These conflicts drove adaptation and strategy changes, leading to outcomes (blue) such as faster service and new team approaches developed through repeated challenges.

Findings

Episode 1: Learning through tool use and shifting roles

Episode 1 captures an early moment in the gameplay where a student named Riley received support from her teammate, the afterschool club teacher, while navigating a task in the game's virtual diner. Riley's object at this point was to prepare a jelly dish for one of the fictional creatures. Riley and the teacher worked together to figure out how to use the tools, manage incomplete information, and respond to changing conditions in the game. This sequence highlights how collaboration can help a student move from uncertainty to more independent and flexible decision-making.

The episode (Excerpt 1 below) begins when the teacher prompted Riley to check the recipe book to find the ingredients needed to make jelly (Turn 1). Riley, unsure about which tool to use, asked, "This is the recipe board?" (Turn 2). Her question showed that she was still unfamiliar with the interface. In response, the teacher

pointed out the correct location of the tool (Turn 3). Riley then admitted, “I don’t know where everything is” (Turn 4), signaling broader uncertainty about her role in the diner task.

Excerpt 1

Turn-by-Turn Interaction Between Teacher and Riley During Gameplay

Turn	Speaker	Talk-In-Interaction	Strategic Function	Activity Theory Element
1	Teacher	Now what are the ingredients? Go to the recipe board and see the ingredients for the jelly.	Referred to tool for decision	Tools
2	Riley	This is the recipe board? ((hovers over data board))	Sought tool confirmation	Tools
3	Teacher	Oh no, no, inside the kitchen yeah.	Clarified tool location	Tools
4	Riley	I don't know where everything is.	Expressed uncertainty	Division of Labor
5	Teacher	I'mma just ask them... ((opens creature history)) what they ate so we could see... nothing... [they are] sending new people.	Retrieved data; interpreted output; Updated team	Tools and Object

Rather than giving Riley step-by-step instructions, the teacher modeled an alternative strategy by checking the “feeding history” of the creature they were serving (Turn 5). When players talked to creatures, they could select different queries that generated data tables. In this case, the teacher selected the rock creature “Bob” and used the query, “What you have eaten?”. This showed Riley another way to gather information. When no feeding data appeared, the teacher explained that she found “nothing… [they are] sending new people” (Turn 5), signaling that the creatures currently seated had not eaten there before, so the group could not rely on past patterns.

From an Activity Theory perspective, this interaction illustrates how tool use and social mediation reorganized the activity system. As Riley learned to navigate the recipe board and creature history tools, the division of labor shifted—she moved from receiving help to actively directing the shared task. The teacher’s guidance functioned as mediated scaffolding, supporting Riley’s ability to connect her actions to the object of serving the creature efficiently. This transformation shows how new forms of agency emerge when participants appropriate tools and strategies within a collaborative context. Rather than simply mastering a game mechanic, Riley’s evolving participation reflected a broader reconfiguration of roles and goals within the system, demonstrating how learning occurs through the coordination of tools, rules, and social interaction.

Episode 2: Competing through tool use and social claims

In Isles of Ilkmaar’s diner space, any team could serve a creature once it was seated, even if another team had prepared the food. This episode shows the first time Cookie and Brownie directly confronted that rule by serving food prepared by others (Excerpt 2 below). Brownie noticed a finished cupcake at the heat lamp, a shared tool, and served it before the other team could act. She announced, “Okay, I got it, got it” (Turn 1). The teacher, playing on the opposing team, immediately contested the action (Turn 2). Brownie reasserted her claim (Turn 3), while Riley questioned what had happened, revealing how quickly access to the tool had shifted (Turn 4). Brownie confirmed ownership through her inventory (Turn 5), and when pressed for proof (Turn 6), she delivered the cupcake to the creature (Turn 7). Cookie reinforced the move with humor, saying, “Bro, all y’all cooking is poor” (Turn 8), shaping how the community interpreted the exchange.

Excerpt 2

Turn-by-Turn Interaction Between Brownie, Riley, Cookie, and Teacher During Competitive Gameplay

Turn	Speaker	Talk-In-Interaction	Strategic Function	Activity Theory Node
1	Brownie	Okay, I got it, got it. ((takes cupcake from heat lamp))	Claimed tool for advantage	Tools
2	Teacher	No, you didn't, no you didn't.	Disputed claim	Rules / Community
3	Brownie	Yes, I did, I got it.	Reasserted action	Tools
4	Riley	Where she…got it? ((checks heat lamp))	Asked for clarification	Community
5	Brownie	It's in my inventory. ((opens inventory))	Confirmed ownership	Rules
6	Teacher	How? Let me see. ((lean over to look at Brownie’s computer))	Demanded proof	Rules / Division of Labor
7	Brownie	Because I took it. ((serves cupcake))	Delivered to prove claim	Division of Labor

 © ISLS

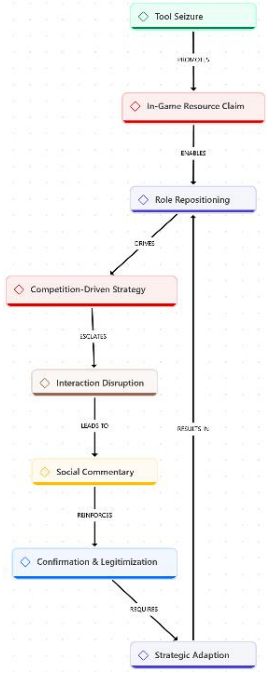
8	Cookie	Bro, all y'all cooking is poor. ((reacts to creature's response))	Critiqued opponents	Community
9	Cookie	((laughs))	Expressed amusement reinforcing rivalry	Community

This episode illustrates how a *tool seizure* triggered a chain of interdependent shifts within the activity system, summarized in Figure 4. As shown in Figure 4, Brownie's grab of the cupcake initiated an *in-game resource claim*, allowing her to move from follower to initiator within her team. That *role repositioning* then drove a *competition-driven strategy* that escalated into disruption and public commentary, eventually requiring *confirmation and legitimation* as others adjusted their responses before the system stabilized. Together, these linked actions show how *tools*, *rules*, and *community* responses interacted dynamically to manage competition in real time. By negotiating what counted as legitimate play and adapting roles under pressure, Brownie and her team transformed moment-to-moment conflicts into socially reinforced strategies.

Figure 4

Sequential Decision Pathway and Network Visualization of Contested Tool Claim in Episode 2

Step	Starting Node	Relationship	Next Node
1	Tool Seizure	Promotes	In-Game Resource Claim
2	In-Game Resource Claim	Enables	Role Repositioning
3	Role Repositioning	Drives	Competition-Driven Strategy
4	Competition-Driven Strategy	Escalates	Interaction Disruption
5	Interaction Disruption	Leads To	Social Commentary
6	Social Commentary	Supports	Confirmation & Legitimization
7	Confirmation & Legitimization	Requires	Strategic Adaptation
8	Strategic Adaptation	Stabilizes	Role Repositioning



(a) (b)

Note. (a) stepwise schematic summarizing Episode 2 decisions. (b) network map of node relationships.

Episode 3: Adapting through conflict and rule enforcement

In one of the more contested diner moments, Riley and the teacher adapted their strategy after Cookie and Brownie repeatedly used the shared icebox for their advantage (Excerpt 3). Riley tried to block competitors by moving all available ingredients into her personal inventory (Turn 1), seizing control of a key tool. Brownie countered by dismissing the move and asserting her team's control over progress (Turn 2), while the teacher asked Riley to redistribute items to keep their team coordinated (Turn 3), invoking the division of labor. Instead, Riley suggested a trade (Turn 4), treating resource exchange as a new rule for fairness. Cookie reinforced her team's dominance with a reminder of their earlier lead (Turn 5), positioning herself within the community hierarchy. As tension grew, the teacher clarified the rule that items had to be placed in the icebox before cooking (Turn 6). Unaware of this constraint, Riley attempted to cook directly from her inventory (Turn 7), prompting the teacher to correct her and reassert procedural rules (Turn 8), explaining the proper steps (Turn 8). Afterwards, Riley and the teacher began timing their actions more carefully, placing ingredients in the icebox and quickly retrieving them to finish tasks while minimizing further interference from the other teams.

Excerpt 3

Turn-by-Turn Interaction of Riley, Brownie, Cookie, and Teacher in Rule Negotiation

Turn	Speaker	Talk-in-Interaction	Strategic Function	Activity Theory Node

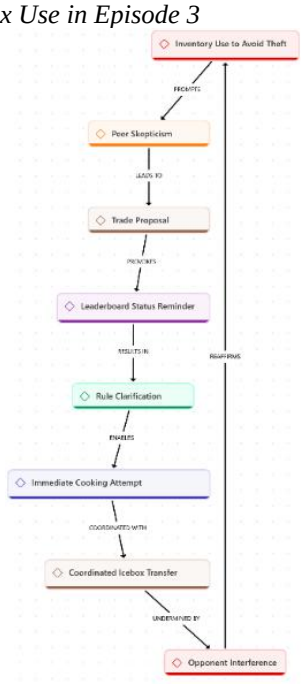
1	Riley	I took everything in the ice box I think	Blocked others by taking items	Tools
2	Brownie	Who's making the food? You. Who's supplying me? You. it doesn't matter	Deflected threat, asserted control	Community
3	Teacher	Okay give it to me	Tried to redistribute items	Division of Labor
4	Riley	I want to trade you I want to trade you everything	Offered item trade	Rules
5	Cookie	Girl you don't know who got the upper [hand]	Claimed dominance	Community
6	Teacher	It still has a piece four at a time ((checking inventory))	Stated inventory rule	Rules
7	Riley	Try to cook real quick	Attempted shortcut cook	Object
8	Teacher	You can't. [Items] have to be in the ice box you cook for ((Explaining the only way to cook is to use the icebox))	Corrected procedure	Rules

The sequence above shows how contested tools and shifting interpretations of rules shaped moment-to-moment strategy, with these dynamics represented in Figure 5, as Riley and the teacher coordinated timing within the division of labor while resisting rival teams. Figure 5 highlights how contested tool use in Episode 3 set off a chain of shifts shaped largely by rules and accountability. Riley's attempt to block competitors by removing all items from the icebox initiated a sequence that quickly drew responses from both teammates and opponents. Brownie dismissed the move, while the teacher stepped in to request redistribution, briefly invoking division of labor to restore balance. The conflict then pivoted around rules: Riley proposed a trade to negotiate fairness, and the teacher clarified the cooking procedure and the requirement that items pass through the icebox. These exchanges reveal how tool use and rules interacted to shape what counted as legitimate action, as determined by the club community, with responsibilities shifting as students sought to secure advantage while staying within the activity system.

Figure 5

Sequential Decision Pathway and Network Visualization of Contested Icebox Use in Episode 3

Step	Starting Node	Relationship	Next Node
1	Inventory Use to Avoid Theft	Prompts	Peer Skepticism
2	Peer Skepticism	Leads To	Trade Proposal
3	Trade Proposal	Provokes	Leaderboard Status Reminder
4	Leaderboard Status Reminder	Results In	Rule Clarification
5	Rule Clarification	Enables	Immediate Cooking Attempt
6	Immediate Cooking Attempt	Coordinated With	Coordinated Icebox Transfer
7	Coordinated Icebox Transfer	Undermined By	Opponent Interference
8	Opponent Interference	Reaffirms	Inventory Use to Avoid Theft



(a) (b)

Note. (a) Stepwise schematic summarizing Episode 3 decisions. (b) network map of node relationships.

Across the three episodes, students' strategic decision-making reflected a learning progression—from initial exploration to contested tool claims and, finally, to rule negotiation under pressure. In Episode 1, Riley relied on the teacher's guidance to interpret tools, gradually developing autonomous problem-solving. In Episode

 © ISLS

2, Brownie and Cookie transformed a conflict over the heat lamp into a competitive learning opportunity, using legitimacy and community feedback to refine their tactics. In Episode 3, Riley and the teacher adapted their use of the icebox, encountering community resistance that prompted clarification and procedural learning. Together, these episodes show collaboration and competition as mechanisms for learning, as reflected in the sequential patterns shown in Figures 4 and 5, with students seizing tools, redefining roles, and testing rules. These evolving interactions reveal how learning emerged through the interdependence of tools, rules, and community, with contradictions in the activity system guiding students' adaptive strategies.

Discussion

This study contributes to CSCL by showing how collaborative and competitive dynamics function as intertwined forms of learning and participation in game-based environments. In the *Isles of Ilkmaar* diner setting, students coordinated responsibilities while monitoring how their group compared against rivals on the leaderboard. This dual focus made decision-making both cooperative and competitive, requiring players to maintain efficiency. Consistent with prior CSCL and game-based learning research on the co-occurrence of collaboration and competition (Qiao et al., 2024; Wang & Huang, 2023), the analysis extends CSCL research by showing that collaboration and competition can be intertwined, prompting participation under layered pressures.

From an Activity Theory perspective, the findings illustrate how contradictions among rules, tools, and roles shaped student choices and adaptations. The diner's tools (recipe log, icebox, heat lamp, and leaderboard) created overlapping demands: rules of efficiency pressed players to keep meals moving, while competitive comparison heightened the urgency of each action. Students' talk about standings, and their adjustments to who served versus who prepared, illustrated how roles were reconfigured as contradictions emerged. This reinforces Engeström's view that contradictions are productive sites of activity, while showing how they operate in fast-paced digital settings where cooperation and rivalry intersect.

The study further contributes to CSCL and the learning sciences by illustrating how micro-level transcripts of gameplay can illuminate the mechanisms through which students balance collective goals with external competition. Methodologically, integrating transcript excerpts with Activity Theory coding demonstrates how fine-grained interaction analysis can reveal strategic decision-making processes. The findings show that students used adaptive, context-sensitive strategies as they negotiated tools, rules, and roles to manage both cooperation and rivalry during gameplay.

Conclusion and future directions

This paper examines how middle school students engaged in strategic decision-making when collaboration and competition were tightly intertwined in an educational gameplay setting. From an Activity Theory perspective, the analysis showed how tools, rules, and shifting responsibilities dynamically shaped student participation.

Several limitations should be noted. While our cases support a broader understanding of the nature of competitive and collaborative learning in CSCL environments, our claims are specific to our setting's activity system. In addition, one team included the teacher as a participant, reflecting an interaction structure with distinct power and knowledge relationships compared to only student-student dynamics (Allen & Hill, 2024), shaping how roles, coordination, and decision-making unfolded during gameplay.

Collaboration and competition did not operate as separate modes of activity; instead, they worked together, each shaping the conditions of the other. Students sustained participation by asserting claims, legitimizing actions, and renegotiating responsibilities, with moments of conflict often serving as catalysts for adaptation rather than breakdowns. This aligns with prior research on productive tension and participation in collaborative learning environments (Engle & Conant, 2002; Engle, 2012). In this way, the analysis shows that cooperative work in game-based environments is inseparable from the competitive pressures that shape it.

These results also build on prior work on gamification and collaborative learning in game-based environments (Qiao et al., 2024) by suggesting that competition can be framed as a productive tension, alongside prior work that has documented its potential negative effects (Cagiltay et al., 2015; ter Vrugte et al., 2015), that pushes students to negotiate fairness, accountability, and efficiency under real-time constraints. When balanced with collaborative structures, such tensions can mirror authentic problem-solving environments where learners must manage shared goals while navigating external pressures. This connects to broader learning sciences perspectives on productive struggle, including productive failure in game contexts (Anderson et al., 2018) and productive disciplinary engagement in collaborative communities (Engle & Conant, 2002; Engle, 2012). Taken together, our study highlights how participation is sustained under layered social and structural demands in game-based learning environments. Future work can examine additional contexts where collaboration and competition coexist, extending understanding of how students adapt, coordinate, and make decisions in complex environments.

References

Allen, C. D., & Hill, K. (2024). "We are all learning here!": Negotiations of power within an intergenerational participatory design research project. In R. Lindgren, T. I. Asino, E. A. Kyza, C. K. Looi, D. T. Keifert, & E. Suárez (Eds.), *Proceedings of the 18th International Conference of the Learning Sciences (ICLS 2024)* (pp. 610–617). International Society of the Learning Sciences.

Almås, H., & Giæver, F. (2024). The emergence of collaboration in serious games: An exploratory study. *International Journal of Serious Games, 11*(3), 89–108. https://doi.org/10.17083/ijsg.v11i3.739

Anderson, C. G., Dalsen, J., Kumar, V., Berland, M., & Steinkuehler, C. (2018). Failing up: How failure in a game environment promotes learning through discourse. *Thinking Skills and Creativity, 30*, 135–144.

Cagiltay, N. E., Ozcelik, E., & Ozcelik, N. S. (2015). The effect of competition on learning in games. *Computers & Education, 87*, 35–41. https://doi.org/10.1016/j.compedu.2015.04.001

Chen, C.-H., & Hwang, G.-J. (2017). Effects of the team competition-based ubiquitous gaming approach on students' interactive patterns, collective efficacy, and awareness of collaboration and communication. *Journal of Educational Technology & Society, 20*(1), 87–98. http://www.jstor.org/stable/jeductechsoci.20.1.87

Danish, J. A., Enyedy, N., Saleh, A., & Humburg, M. (2020). Learning in embodied activity framework: A sociocultural framework for embodied cognition. *International Journal of Computer-Supported Collaborative Learning, 15*(1).

Derry, S. J., Pea, R. D., Barron, B., Engle, R. A., Erickson, F., Goldman, R., … Sherin, B. L. (2010). Conducting video research in the learning sciences: Guidance on selection, analysis, technology, and ethics. *Journal of the Learning Sciences, 19*(1), 3–53. https://doi.org/10.1080/10508400903452884

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: Defining "gamification." In *Proceedings of the 15th International Academic MindTrek Conference: Envisioning future media environments* (pp. 9–15). https://doi.org/10.1145/2181037.2181040

Engle, R. A., & Conant, F. R. (2002). Guiding principles for fostering productive disciplinary engagement: Explaining an emergent argument in a community of learners classroom. *Cognition and Instruction, 20*(4), 399–483.

Engle, R. A. (2012). The productive disciplinary engagement framework: Origins, key concepts, and developments. In D. Y. Dai (Ed.), *Design research on learning and thinking in educational settings (pp. 161–200)*. Routledge.

Fischer, F., Kollar, I., Stegmann, K., & Wecker, C. (2013). Toward a script theory of guidance in computer-supported collaborative learning. *Educational Psychologist, 48*(1), 56–66. https://doi.org/10.1080/00461520.2012.748005

Kim, B., & Ho, W. (2018). Emergent social practices of Singapore students: The role of laughter and humour in educational gameplay. *International Journal of Child-Computer Interaction, 16*, 85–99. https://doi.org/10.1016/j.ijcci.2018.01.001

Miller, M., & Hadwin, A. F. (2015). Scripting and awareness tools for regulating collaborative learning: Changing the landscape of support in CSCL. *Computers in Human Behavior, 52*, 573–588. https://doi.org/10.1016/j.chb.2015.01.050

Moon, K., & Ke, F. (2020). Exploring the relationships among middle school students' peer interactions, task efficiency, and learning engagement in game-based math learning. *Simulation & Gaming, 51*(3), 311–336. https://doi.org/10.1177/1046878120907940

Pareto, L., Haake, M., Lindström, P., Sjödén, B., & Gulz, A. (2012). A teachable-agent-based game affording collaboration and competition: Evaluating math comprehension and motivation. *Educational Technology Research and Development, 60*(5), 723–751. https://doi.org/10.1007/s11423-012-9246-5

Pervez, A., & Orchard, S. (2025). Reimagining curricula: A case study of a game-based approach to fostering prosociality and cross-subject learning in primary schools. *Education 3-13, 53*(7), 1142–1161. https://doi.org/10.1080/03004279.2025.2524480

Plass, J. L., O'Keefe, P. A., Homer, B. D., Case, J., Hayward, E. O., Stein, M., & Perlin, K. (2013). The impact of individual, competitive, and collaborative mathematics game play on learning, performance, and motivation. *Journal of Educational Psychology, 105*(4), 1050–1066. https://doi.org/10.1037/a0032688

Qiao, S., Yeung, S. S., Shen, X. A., Leung, K. L., Ng, T. K., & Chu, S. K. W. (2024). How competitive, cooperative, and collaborative gamification impacts student learning and engagement. *Language Learning & Technology, 28*(1), 1–19. https://doi.org/10.64152/10125/73546

Radke, S. C., Hardy, L., & Kahn, J. B. (2024). Youth identity enactments through storytelling during co-design of an educational virtual game world. In R. Lindgren, T. I. Asino, E. A. Kyza, C. K. Looi, D. T. Keifert, & E. Suárez (Eds.), *Proceedings of the 18th International Conference of the Learning Sciences (ICLS*

2024) (pp. 1538–1541). International Society of the Learning Sciences. https://doi.org/10.22318/icls2024.703626

Saldaña, J. (2016). *The coding manual for qualitative researchers.* Sage.

Suthers, D. D. (2006). Technology affordances for intersubjective meaning making: A research agenda for CSCL. *Computer-Supported Collaborative Learning, 1*(3), 315–337.

ter Vrugte, J., Bastiaens, T., & van Riesen, S. A. (2015). Effects of collaboration and competition in game-based learning on mathematics performance of prevocational students. *Computers & Education, 89*, 42–52. https://doi.org/10.1016/j.compedu.2015.08.010

Van der Meij, H., de Jong, T., & Caski, W. (2011). Learning from games: Does collaboration help? *British Journal of Educational Technology, 42*(4), 655–664. https://doi.org/10.1111/j.1467-8535.2010.01067.x

Wang, D.-C., & Huang, Y.-M. (2023). Exploring the influence of competition and collaboration on learning performance in digital game-based learning. *Educational Technology Research and Development, 71*(4), 1547–1565. https://doi.org/10.1007/s11423-023-10247-8

Xia, Y., Yan, S., Jiang, M., & Brown, Z. (2025). Supporting learning in synchronous collaborative game design in virtual worlds: A synergy between technological and pedagogical considerations. *Computers & Education: X Reality, 7*, Article 100110. https://doi.org/10.1016/j.cexr.2025.100110

Yu, F.-Y. (2001). Competition within computer-assisted cooperative learning environments: Cognitive, affective, and social outcomes. *Journal of Educational Computing Research, 24*(2), 99–117. https://doi.org/10.2190/3U7R-DCD5-F6T1-QKRJ

Acknowledgments

This work was supported by NSF Award #2214516.

International Society of the Learning Sciences

Exploring Variations in Teacher Modelling and Scaffolding Through Chatbot Interactions in Middle School Classrooms

Henrik Tjønn, Sten Ludvigsen, and Anders Mørch
henrik.tjonn@iped.uio.no, sten.ludvigsen@iped.uio.no, anders.morch@iped.uio.no
School of Education, University of Oslo

Abstract: In this paper, we develop new insights into how teachers can model, scaffold and coach students when using tailored chatbots in a school subject. Our interaction analysis, based on a cognitive apprenticeship framework, shows how teachers affirm student statements regarding the chatbot's lack of knowledge structures. We identify a few instances of teacher elaboration on the chatbots' subject-specific knowledge structures during classroom interactions. Teachers mainly provide concrete scaffolding support for the tasks and confirm students' questions and claims about inaccuracies in chatbot outputs. Our analysis also indicates that the balance between modelling and scaffolding depends on how specific sequences in teaching and learning activities are executed.

Background

The emergence of generative artificial intelligence (genAI) presents significant opportunities and challenges for classroom research, particularly concerning contextualization, knowledge, and agency (Bastani et al., 2025; Klar, 2025; Warschauer et al., 2023). Researchers in AI and education highlight the importance of both students and teachers understanding the functions of large language models (LLMs) and their potential to enhance specific learning processes. Moreover, it is vital to establish new collaborative repertoire and roles among AI, students, and teachers (Tjønn et al., 2025).

Partnerships with schools to support development in the classroom is a key issue within the field of CSCL. In the research project "Learning in the Age of Algorithms" (LAA), we have designed chatbots with subject-specific and generative capabilities for use in middle schools together with school developers and teachers in a Norwegian municipality. It has been necessary to develop new designs for teaching in classrooms that incorporate chatbots, creating what we refer to as local cultural fit (Ludvigsen & Mørch, 2003). In our research, we aim to describe and understand how chatbots are utilized in middle school classrooms with students aged 10 to 16, including the teacher's instructions. To analyse teacher scaffolding and modelling of interactions with chatbots, we pose the following research question in this paper (*RQ1*): *What characterizes teachers' scaffolding and modelling of students' chatbot use in two Norwegian middle school classrooms, as analyzed through the cognitive apprenticeship framework?*

RQ1 guided our analysis of how teachers model their interactions with chatbots for 9th grade students to observe. We used a conceptual framework (analytic stance) to anchor the analysis. This framework incorporates key concepts in collaborative learning that capture the intersection of social and cognitive processes. Our analysis builds upon existing literature regarding the use of AI tools in K-12 classrooms.

Literature review

In this brief review, we emphasize studies that provide insights into the early phases of AI in education, as well as recent studies utilizing various forms of genAI, including both quantitative and qualitative research. We argue that in this new phase, it is essential to examine different research designs to understand how students interact with and learn from genAI.

Interaction with different AI tools in the classroom has been explored in educational research, particularly AI in education, computers and education, including Computer-supported collaborative learning (CSCL). A classical CSCL study by Baker and Bielaczyc (1995) investigated how students aged 16–17 engaged with the CHENE system, an AI platform designed for scientific modelling in physics. It aimed to foster co-construction of explanations using scientific vocabulary between student partners in interaction. The study revealed missed opportunities; only one student engaged in higher-level reasoning while their partner did not, limiting their chances for deeper understanding. In contrast, Suthers et al. (1997) examined the networked software system, Belvedere, which used AI coaches to support groups of 2-6 students. Observations showed that students found the activities engaging and teachers were satisfied with their training. The authors asserted that the success of CSCL technology depends on its acceptance by teachers (Suthers et al., 1997). However, effective dialogues in computational learning design cannot be assumed, and cognitive challenges persist even with

 © ISLS

technological advancements (Baker et al., 2021). These limitations highlight the necessity for further exploration of teachers' roles with genAI tools in educational contexts, as genAI, in contrast to systems like CHENE/Belvedere, do not function as knowledge systems (Holmes & Tuomi, 2022).

Borge et al. (2024) examined how interactions with ChatGPT 3.5 could foster new learning environments that support socio-metacognition. Their study revealed promising results from an experienced CSCL researcher interacting with ChatGPT, highlighting a pattern of both domain-specific and generic expertise. The authors suggested that genAI is particularly effective when students are shown expert interactions with chatbots (modelling). However, they emphasized the need for further exploration in natural classroom settings and adapting models for student use across various levels and subjects. Chen et al. (2023) illustrated differences in student experiences with generative AI through their study on high school students (n=10). Interviews revealed that students found ChatGPT helpful for information retrieval and completing monotonous tasks but had mixed views on its potential to enhance more advanced knowledge building, as well as varying levels of critical thinking toward its use. While both studies imply the importance of modelling and support, neither study consists of explicit examples of teacher modelling/scaffolding.

We have identified relevant experimental studies of the use of genAI in K-12 classrooms as part of a general research exploration stage. Klar (2025) examined how 9th and 10th graders (n=106) interacted with a generic version of ChatGPT 3.5. The study compared pre- and post-tests of an experimental group receiving brief guidance on adaptation prompting with a control group that did not. Students found the chatbot easy to use, while the control group underutilized its adaptation capabilities. The experimental group used adaptation prompts twice as often (M=6.0) as the control group (M=2.5), indicating different interaction patterns. Both groups, however, showed similar knowledge gains and high satisfaction levels. The study is based on a 50-minute session (Klar, 2025). In contrast, Lademann et al. (2025) conducted a randomized control study with 6th grade students (n=214) using a domain-specific, custom chatbot based on ChatGPT 4.0. The results demonstrated that AI-generated material significantly impacted students' situational interest and self-efficacy while reducing cognitive load. However, its impact on learning outcomes was not significant.

Recent CSCL studies

Research on genAI interaction in K-12 classrooms is emerging within CSCL. Ohsaki and Matsukawa (2025) analyzed high school students' (n=280) interactions with genAI, finding that the chatbot frequently provided advice to science project groups without explicit requests. Their analysis revealed significant differences in message lengths (436.3 characters for chatbots vs. 54.8 for user messages), suggesting limited engagement from students. Xin et al. (2025) introduced CocoNote, a genAI-driven platform that enhances students' socially shared regulation (SSRL) in collaborative environments. A quasi-experimental study with 6th graders (n=78) showed significant improvements in SSRL capabilities, supported by positive feedback from five interviewed students. Ojeda-Ramirez et al. (2025) explored how elementary and middle school students (n=30) aged 6-12 engaged with generative AI in storytelling (Story AI), revealing that students balance AI contributions with their own creativity and reflect on its role in their narratives.

Our literature review reveals a lack of studies on the interaction between students and teachers in naturalistic settings (classrooms). Few genAI studies provide detailed descriptions of chatbot design, teacher and student perceptions, or analysis of the actual partnerships and teacher modelling occurring in K-12 classrooms during experiments. Additionally, limited research has explored instructional designs that integrate tailored genAI into learning activities. Consequently, current research does not adequately address how genAI technology can be combined with teacher instruction, ongoing (e.g. design-based) research, diverse resources, and interactions among teachers, students, and peers. This gap is what we aim to address by investigating RQ1.

Analytic stance

Modelling and scaffolding are key concepts in our analysis. Both are central components of the approach known as Cognitive Apprenticeship (CA) (Collins et al., 1991; Collins & Kapur, 2014), applied in classrooms. Collins et al. (1991) argue that insufficient attention is often given to the reasoning strategies experts use to acquire and apply knowledge for complex real-world tasks. Consequently, the conceptual and problem-solving knowledge students gain in school often goes unused; they struggle with problems that deviate from established patterns and neglect available resources due to a lack of effective models. To improve students' skills, it is essential to understand expert practices and develop appropriate learning methods. In the CA approach, the actors' thinking must be made visible for both teachers and students, whether in reading, writing, or problem-solving. Additionally, CA seeks to contextualize abstract tasks in ways which are meaningful to students and introduces a range of tasks from systematic to varied, encouraging students to reflect on and articulate common characteristics. The goal is to help students generalize the skill, recognize its relevance, and transfer it to new situations

independently. Collins et al. (1991) have developed a framework outlining principles for designing cognitive apprenticeship environments. The framework in Table 1 consists of the following dimensions (D) and characteristics (C):

Table 1
Designing For Cognitive Apprenticeship (CA) Environments (Collins et al., 1991)

	1. Content (types of knowledge required for expertise)	**2. Method** (ways to promote the development of expertise)	**3. Sequencing** (keys to ordering learning activities)	**4. Sociology** (social characteristics of learning environments)
D:				
C:	Domain knowledge (subject specific concepts, facts).	Modelling (teacher performs a task so students can observe).	Global before local skills (focus on conceptualizing the whole task).	Situated learning (students learn in the context of working on realistic tasks).
	Heuristic strategies (generally applicable techniques for completing tasks)	Coaching (teacher observes and facilitates while students perform a task).	Increasing complexity (meaningful tasks gradually increasing in difficulty).	Community of practice (communication about different ways to accomplish meaningful tasks)
	Control strategies (general approaches for directing one's solution process)	Scaffolding (teacher provides supports to help the student perform a task)	Increasing diversity (practice in a variety of situations to emphasize broad application)	Intrinsic motivation (personal goals to seek skills and solutions) and co-operation

Dimensions (D) and characteristics (C) of a modified version of *Principles for designing cognitive apprenticeship environments* (Collins et al., 1991). We primarily utilize the defined characteristics of the Content and Methods dimensions, as these serve as the central focus of the empirical analysis.

In addition to the descriptions in Table 1, modelling involves students constructing conceptual models of the processes needed to complete a task. Coaching also entails providing hints and feedback related to specific issues that arise as students work to accomplish a task (Collins & Kapur, 2014). Scaffolding refers to the support a teacher provides to help students complete a task (Wood, Bruner & Ross, 1976). An important distinction between coaching and scaffolding is that scaffolding more specifically pertains to the support given to the learner and can take the form of suggestions or direct help. The timing of this support is critical: it can be offered initially and then gradually fade as students gain independence, or it can be provided after the learner encounters an impasse or fails to complete a task (Collins & Kapur, 2014). Additionally, the CA framework consists of additional characteristics of the Method dimension: articulation (the teacher encourages students to articulate their knowledge, problem-solving processes or reasoning), reflection (the teacher enables students to compare their performance with that of their peers or an expert) and exploration (the teacher invites students to formulate and solve their own problems) (Collins et al., 1991, Collins & Kapur, 2014).

Like the previously mentioned genAI studies, interventions based on CA in K-12 classrooms have also been investigated in recent studies. Wang et al. (2024) conducted a quasi-experimental study with high school students (n=150) that found cognitive apprenticeship models in math outperformed conventional teacher-centred instruction (Wang et al., 2024). Similarly, De La Paz et al. (2024) examined 8th grade students (n=110) argumentative writing in science. The results show that an intervention based on a cognitive apprenticeship approach helped students develop skills needed to improve their written scientific arguments (De La Paz et al., 2024).

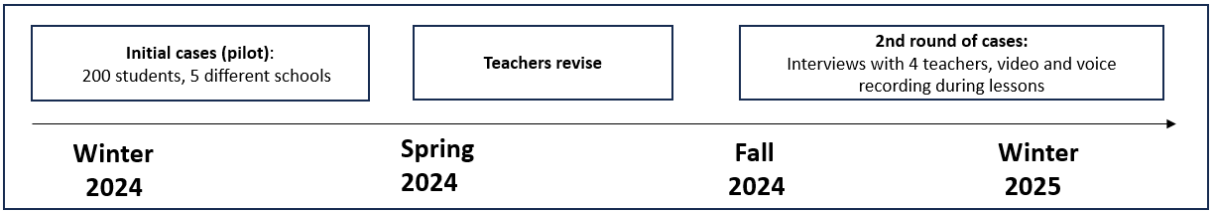
To understand how classroom dialogue develops knowledge and understanding, we must examine the temporal relationship between the organization of teaching and learning as sequences of lessons and activities and how this is enacted through conversation (Mercer, 2019, pp. 190–191). Our analytical approach draws on key Cognitive Apprenticeship (CA) concepts in Table 1, focusing on how teacher modelling, coaching and scaffolding of chatbot interaction are manifested in classroom dialogue (Bakhtin, 1986; Greeno & Van de Sande, 2007; Tabak & Kyza, 2018).

Methods

The LAA project is framed as Design-Based Research (DBR), which is characterized by the design of interventions (McKenney & Reeves, 2018). Our project partner, a municipality in Norway, designed various chatbots based on ChatGPT 4.0 (SchoolGPT). Our design principles are based on classrooms as communities where situated learning occurs and emphasizes that learning trajectories require specific types of sequencing. By using system prompts, pedagogical advisors customized ChatGPT 4.0 to suit the language and knowledge of students aged 10 to 16. The chatbots were also tailored to specific subjects (e.g., science, history) by integrating the national curriculum (e.g. curricular documents) into the design. To fine-tune the design, designers gradually refined the system prompts based on feedback from participating researchers and teachers (Mørch et al., 2025). The customized SchoolGPT can be described as a set of distributed prompts in specific sequences. A key aim of this design work is to explore how chatbots can facilitate critical thinking in school subjects through students' subject-specific interactions. (Tjønn et al., 2025). From Winter 2024 to Winter 2025, researchers from LAA observed how students interacted with the adapted chatbots during school lessons. The timeline for this period is illustrated in Figure 1.

Figure 1

LAA Project Timeline, 2024–2025



The complete data corpus from the second round of cases includes classroom recordings of two observed teachers over a period of 4-5 weeks. We video recorded the teachers' instructions and student pairs' interactions with chatbots. The corpus also consists of audio recordings of interviews with 7-8 focus students per class and two additional teachers, as well as written lesson plans and screen recordings of students' work. We utilized an institutionally approved AI tool to transcribe all the classroom audio recordings. The first author then listened to the recordings and manually adjusted the transcriptions for lessons involving student interactions with chatbots. Additionally, all three authors reviewed the video recordings of teacher instruction and student interactions together, reaching a consensus on the descriptions presented as "Context" in the excerpts in Table 2-6. The data chosen for presentation demonstrate the central phenomenon.

The following analysis employs an interaction analysis approach (Jordan & Henderson, 1995) and focuses on data from the SchoolGPT-related interactions that occurred during the lessons of the two observed teachers, "Mike" and "Ted". The interaction analysis uses the CA framework as a heuristic strategy.

Classroom interaction with SchoolGPT

We present excerpts from video and audio recordings of observed teachers' classrooms, drawn from lessons where students interact with SchoolGPT. Pauses over 2 seconds are marked "(…)" in the transcriptions. Our interpretation of the primary CA characteristic types will be described in the paragraphs following the excerpt tables. We also describe the time frame and pedagogical context of students' chatbot use.

Teacher 1: "Mike" – examples of scaffolding and modelling

Mike teaches science at a middle school in the municipality. His class previously worked with chatbots as part of the LAA project pilot. In the first observed lesson, Mike initially introduces a project in which his 9th grade students will use SchoolGPT to prepare for a role play by asking the chatbot to respond as a famous environmental scientist. The chatbot's output will be used in an assessed oral exercise at the end of the period. Students are divided into groups of four, with two students paired for each scientist (co-operation). They are given the following initial prompt from the teacher as a sequential starting point in their interaction with SchoolGPT:

Teacher:	Hello. I wish to interview [scientist name], who's an acclaimed environmental scientist known for views on greenhouse effect and global warming. Instead of me interviewing the person, I would like you to pretend like you are this person when I ask you some questions?

 © ISLS

Students are told verbally to ask the chatbot for biographical facts and details about the greenhouse effect. As a next sequential step, they will check their responses using other digital resources. During his introduction, Mike emphasizes the importance of critical thinking in students' assessment when interacting with SchoolGPT:

> [Mike]: Artificial intelligence, it has a lot of information and provides many answers, but sometimes it can hallucinate... make things up, right? So we need to ensure that you are receiving accurate information.

The teacher pastes this prompt into a new conversation with SchoolGPT on a screen in front of the class (modelling chatbot interaction, global before local skills). Students work in pairs at one personal computer to read the chatbot's answers and record responses in a Google Docs document table, pasting screenshots of the conversation at the bottom. The teacher then shows students how to copy and paste these screenshots (modelling). Transcriptions of student-student-teacher interactions are presented in Table 2.

Table 2
Mike's Interaction with S1 and S2

#	Actor	Utterance	Context
1	S1	Everything he [the chatbot] says here is wrong.	*The teacher passes by*
2	S2	Is Richard Lindzen Norwegian?	*The student asks Mike*
3	S1	"He was born in a small town in Norway."	*S1 reads out loud*
4	M	Oh:: That is.. completely wrong.. (…)	*The students laugh*
5	M	Lindz**e**n.. Maybe an e instead of an o? Lindzon is kind of... Norwegian.	*The teacher reads*
6	S1	Oh yes, sorry..	*Confirmation*
7	M	Use 'e'.	*Correct spelling*
8	S1	And then z..	*S1 begins to write*
9	M	And z, right? Like that, yes.. And then the next question, about background.. Now it's starting to come together! (…) But.. check that information, guys. Go on Wikipedia, for example. Write about him and compare the two answers.	*Teacher points to the screen*

Here, we observe the conjunctions of S1 and S2 seeking Mike's support by stating that a knowledge claim in the chatbot output was incorrect. The confirmation of S1's statement in line #4 functions as coaching of students' interaction with SchoolGPT. After Mike's problem-solving utterances of scaffolding (#5,7), strategies (#9), and gesture enactment like pointing, however, the students articulate that their own input led to the incorrect output. Additionally, he offers the students heuristic and control strategies in #9 by outlining next steps they should take to verify the information provided by SchoolGPT. This same sequential pattern is evident in Mike's active interaction with S3 while the student is collaborating with S4, as Mike writes the input for them in Table 3 (modelling chatbot interaction).

Table 3
Mike's Interaction with S3

#	Actor	Utterance	Context
1	S3	Eh:: It says here... "I don't have knowledge of"...	*S3 read aloud*
2	M	Adjust... adjust the prompt. Change the prompt. (…)	*Mike grabs the computer*
3	M	What should I write? (…)	*Mike asks the students*
4	M	"I... know... that... but can you pretend?" Like this. Now it's started. Now it wants to.	*Teacher writes on the student's computer*
5	S3	But should I just ask... What is the background?	*Explaining action*
6	M	Yes, and then check online if it's correct. Before you move on to the next question.	*Confirming and give advice*

 © ISLS

In this interaction sequence, Mike takes control of the student's computer while asking S3 to articulate what he should write in line #3. Mike then models his reasoning strategies while interacting with SchoolGPT by writing on the computer as students observe him in line #4 (modelling). The students' observation of Mike's enactment of chatbot interaction distinguishes his modelling of interaction from the scaffolding he provides to S1 and S2 in table 2. Like his engagement with S1 and S2, he also explains the strategies S3/S4 should use in their upcoming problem-solving tasks (control strategies). In the following interaction with S5 working with S6 in Table 4, he takes control of their computer but does not write for them.

Table 4
Mike's Interaction with S5

#	Actor	Utterance	Context
1	S5	What do we do if... SchoolGPT is... wrong? It says... everything is correct... except that she was born in 1956 because other places say she was born in 1953.	*Asking Mike*
2	M	Yes. But the rest...	*Confirmation*
3	S5	The rest is correct.	*Confirmation*
4	M	The rest is correct. But then it's not that bad. Let's move on to the next question.	*Teacher turns the computer towards himself*
5	S5	Okay.	*Confirmation*
6	M	But make sure that... (..) As you get responses going forward... make sure you continue in that direction.	*Teacher looks at S5.*
7	S5	Okay.	*Confirmation*
8	M	Paste in and ask the next question. The most important thing for you, girls, is here. When you write here, right?	*Teacher switches windows*

In this sequence, students refer to the chatbot's output and explain its errors through comparison. Students and Mike alternate in questioning and evaluating SchoolGPT's responses, as Mike remarks at line #2 (coaching). In line #4, he prompts the students to advance on to the next question (control strategies, scaffolding). Finally, in line #8, Mike stresses that these inaccuracies are secondary to other task aspects (heuristic/control strategies, scaffolding, coaching, increasing complexity). Mike also models problem solving he is alluding to by switching websites/tabs on the students' computer while the students observe him as he navigates between the digital tools. Toward the end of the lesson, he summarizes the students' work in a whole-class dialogue.

Teacher 2: "Ted" – modelling activities in whole-class discussions

Ted teaches social science and English at another middle school in the municipality. His class was observed during the initial testing; thus his 9th grade students have prior experience using chatbots. In the first observed lesson, one of Ted's initial sequential actions is to display the subject plan and learning goals for the period on a large screen:

> Learn to be critical of sources. Consider advantages and disadvantages of different kinds of sources. Obtain a general overview of the conflict known as the First World War.

The students will work with both analoge and digital sources during the period, including SchoolGPT. During the fifth lesson, Ted facilitates the students' use of SchoolGPT as a source during his lesson introduction:

> [Ted]: Today, we are going to do the same thing, except that... Today we will use SchoolGPT. Or AI. Artificial intelligence. So, before we get started, I want you to talk a little among yourselves about whether you remember... what is AI? What is SchoolGPT? If you recall this? Also, if you remember any tips and tricks for how to get better answers when using it.

In a subsequent whole-class dialogue, students suggest making the content easier to understand by giving SchoolGPT context and boundaries and by checking information with other sources (community of practice). The student task is to identify two significant events from World War I. Students are instructed to conduct a source

critical analysis of the chatbot's output by comparing the SchoolGPT responses with those from other digital sources. Students will work in pairs (co-operation), sharing one computer based on the teacher's starting prompt: "Hi, I'm in 9th grade and I wonder if you can tell me about two important events from World War I".
Before the students work together in pair (co-operation), Ted models his interaction with SchoolGPT on a screen in front of the class. Like Mike, Ted's modelling can be interpreted as a designed sequential step prior to the students working together. Most student-student-teacher interactions during the rest of the lesson concern computer power supply and technical difficulties, but his chatbot-specific conversations with selected students were video-recorded and will be the focus for our analysis. Transcriptions of these interactions are in Table 5, beginning with Ted's engagement with S7 and S8.

Table 5
Ted's Engagement with S7/S8

#	Actor	Utterance	Context
1	S7	Yes, but... look here. Scroll down. Look, there are the sources.	*Direct the attention to Ted*
2	T	"As AI, I do not have direct access to specific sources or databases, and I cannot quote specific books or articles."	*Teacher reads aloud*
3	S8	Ted? I asked what websites, and then it said this.	*S2 points at the screen*
4	T	Yes, that [URLs] is just general domain websites. So, you don't really get... specific sources for what it has written.	*Explain what knowledge can be found in the source*
5	S7	No.	*Confirmation*

Students refer to the chatbot's claims about specific websites. After reading SchoolGPT's output out loud in line #2, Ted explains that chatbots do not reference specific webpages (control strategies, coaching), but the knowledge structure of LLMs is not articulated when he discusses SchoolGPT citing general-domain websites. In the lesson's closing minutes, Ted asks about students' experiences with SchoolGPT in a whole-class discussion; S9 and S10 participate in Table 6:

Table 6
Ted's Engagement With S9/S10

#	Actor	Utterance	Context
1	T	Have you received any answer from the bot... regarding which sources it uses? [S9]?	*Ask question*
2	S9	Eh:: It said that.. "as AI, I cannot provide direct sources".. .... or the internet. That it bases its information on generalized knowledge and such.	*S9 reads aloud.*
3	T	So it uses generalized knowledge...	*Ask more specific question*
4	S9	Yes, it doesn't want to give us...	*Explain the response of the bot*
5	T	No, it doesn't want to provide the specific sources? [S10]?	*Confirmation*
6	S10	It said that it hadn't used specific sources to give the answers but bases the information on general knowledge about World War I.	*Describes the limitation of the bot*
7	T	Generalized knowledge about World War I.. Here... I also asked the same. (…) So it probably says that it.. it cannot provide.. specific sources.. for the time being.	*Ted shows a screenshot from his own interaction*

Initially, Ted facilitates a whole-class assessment of SchoolGPT's output by questioning the traits of ChatGPT's output, repeating the observed student responses in lines #3 and #5 (coaching, articulation). After S10's claim about the chatbot in line #6, Ted demonstrates his previous interaction with SchoolGPT on the screen in line #7 while the students observe him (modelling). Additionally, he makes general remarks about other strategies students may use when interacting with SchoolGPT (heuristic, control strategies).

Discussion

In the following, we will address RQ1: What characterizes teachers' scaffolding and modelling of students' chatbot use in two Norwegian middle school classrooms, as analyzed through the cognitive apprenticeship framework? We interpret the excerpts as showing that both Ted and Mike display several characteristics of different CA dimensions that support students' interactions with chatbots. Consistent with studies of expert and novice chatbot use (e.g., Borge et al., 2024; Chen et al., 2023), notable differences appear in how teachers articulate their rationale for writing inputs and interpreting outputs. Both teachers make their thinking visible by identifying, demonstrating, and correcting issues in students' prompts during whole-class and teacher–student interactions. The CA framework's dimensions (Collins et al., 1991) of coaching, modelling, scaffolding, and problem-solving strategies are relevant in the analysed interactions. In the literature review, some K-12 studies report promising results when students interact with genAI (Chen et al., 2023; Lademann et al., 2025; Xin et al., 2025; Ojeda-Ramirez et al., 2025). As both literature (e.g., Baker et al., 2021; Borge et al., 2024; Klar et al., 2025) and the excerpts demonstrate, novice students may require teacher support when interacting with genAI tools.

Our contribution to the CSCL field is the identification of distributed scaffolding prompts, namely how the teachers share their initial guiding prompts before students engage with the chatbots. In the CA framework, scaffolding denotes the support a teacher provides to help students perform a task (Collins et al., 1991). The scaffolding prompts used by the teachers do not necessarily require continuous teacher intervention during exchanges among students and chatbots, neither must they include demonstrations of chatbot interaction; rather, they could serve as an essential starting point for students' subsequent interaction with the chatbots. We therefore suggest that teachers, as Ted and Mike do, ahead of time design and distribute subject-specific scaffolding prompts before engaging in modelling, coaching, or other kinds of scaffolding in their classrooms.

The teachers' enactment of the Content dimension in the CA framework is an interesting finding. As they model and scaffold chatbot interaction, we find instances where incorrect knowledge claims are neither dealt with nor elaborated upon, or where explanations for SchoolGPT's inaccuracies remain implicit. For example, in line #1 of Table 4, S5 independently flags a partly incorrect claim by referencing other sources. Mike subsequently prompts them to disregard the false birth year, deeming it less central to the task. In Table 5, S7 observes that SchoolGPT cannot provide concrete sources, and Ted confirms but does not explain the model's underlying knowledge structures or limits. In Table 6, Ted tends to confirm students' statements instead of elaborating on how the system generates answers or why it should not be treated as an epistemic authority. Both teachers explicitly emphasize critical thinking about SchoolGPT's output in their introductions, and coaching is an important CA characteristic in these lessons (Collins & Kapur, 2014). However, relying primarily on coaching by confirming student claims without concrete scaffolding or modelling support, could represent missed opportunities to develop students' understanding of the technology and content's epistemic status.

Conclusions and directions for further work

Through their explicit learning goals, initial scaffolding prompts, and whole-class discussions, both teachers highlight and somewhat problematize potential inaccuracies in the domain-specific knowledge presented by SchoolGPT. By questioning and adjusting students' chatbot inputs and by making their reasoning visible when addressing chatbot output, Ted and Mike both scaffold and model how to treat SchoolGPT as an interactive classroom resource. This kind of support could be vital when students interact with genAI tools as emphasized by previous research (e.g., Baker et al., 2021; Borge et al., 2024; Klar et al., 2025).

Beyond providing distributed scaffolding prompts that can support students over a sequence of actions, the teachers demonstrate practical preparation to provide various supports that may occur during lessons. This could include offering explicit explanations for why SchoolGPT and other tools built on LLMs can produce inaccurate information, as these models cannot list the sources for the output. In collaboration with our project partners, we will continue to explore additional support for teachers integrating genAI into instructional design, including further refinement of SchoolGPT aimed at reducing the incidence of inaccurate outputs. Additionally, we will emphasize another important finding, namely how the integration of chatbots into specific content-generating sequences will depend on the lesson design. Overall, our study shows that students need to be taught to always think critically when interacting with chatbots, as the information provided may be incorrect or imprecise.

References

Baker, M. J., & Bielaczyc, K. (1995). Missed opportunities for learning in collaborative problem-solving interactions. In *Proceedings of the AI-ED 95-World Conference on Artificial Intelligence in Education* (pp. 210-217).

Baker, M. J., Schwarz, B. B., & Ludvigsen, S. R. (2021). Educational dialogues and computer supported collaborative learning: Critical analysis and research perspectives. *International Journal of Computer- Supported Collaborative Learning*, *16*, 583–604.

Bakhtin, M. M. (1986). *Speech genres and other late essays*. University of Texas Press.
Borge, M., Smith, B. K., & Aldemir, T. (2024). Using generative AI as a simulation to support higher-order thinking. *International Journal of Computer-Supported Collaborative Learning, 54*(1).
Chen, B., Zhu, X., & Díaz del Castillo H., F. (2023). Integrating generative AI in knowledge building. *Computers and Education. Artificial Intelligence, 5*.
Collins, A., Brown, J. S., & Holum, A. (1991). Cognitive apprenticeship: Making thinking visible. *Am. Educator, 15*(3), 6-11.
Collins, A., & Kapur, M. (2014). Cognitive Apprenticeship. In R. K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (pp. 109–127). Cambridge: Cambridge University Press.
De La Paz, S., Butler, C., Levin, D. M., & Felton, M. K. (2024). Effects of a Cognitive Apprenticeship on Transfer of Argumentative Writing in Middle School Science. *Learning Disability Quarterly, 47*(2), 70–83.
Greeno, J. G., & van de Sande, C. (2007). Perspectival Understanding of Conceptions and Conceptual Growth in Interaction. *Educational Psychologist, 42*(1), 9–23.
Holmes, W., & Tuomi, I. (2022). State of the art and practice in AI in education. *European J. of Education, 57*(4), 542–570.
Jordan, B., & Henderson, A. (1995). Interaction Analysis: Foundations and Practice. *The Journal of the Learning Sciences, 4*(1), 39–103.
Klar, M. (2025). Using ChatGPT is easy, using it effectively is tough? A mixed methods study on K-12 students' perceptions, interaction patterns, and support for learning with generative AI chatbots. *Smart Learning Environments, 12*(1), 19–32.
Lademann, J., Henze, J., & Becker-Genschow, S. (2025). Augmenting learning environments using AI custom chatbots: Effects on learning performance, cognitive load, and affective variables. *Physical Review. Physics Education Research, 21*(1).
Ludvigsen, S., & Mørch, A. (2003). Categorisation in Knowledge Building: Task specific argumentation in a co-located CSCL environment. In B. Wasson, S. Ludvigsen, & U. Hoppe (Eds.), *Designing for Change in Networked Learning Environments* (pp. 67–76). Springer Netherlands.McKenney, S., & Reeves, T. C. (2018). *Conducting Educational Design Research* (2nd ed.). Routledge.
Mercer, N. (2019). *Language and the joint creation of knowledge: the selected works of Neil Mercer*. Routledge.
Mørch, A. I., Ludvigsen, S., & Gilje, Ø. (2025). Teachers as End-User Developers: Two Case Studies of Adapting Language Models for Education. *CEUR Workshop Proceedings, 3978*.
Ohsaki, A., & Matsukawa, H. (2025). Interaction Analysis Between Generative Ai and High School Students in Scientific Projects. In Oshima, J., Chen, B., Vogel, F., & Järvelä, J. (Eds.), *Proceedings of the 18th International Conference on Computer-Supported Collaborative Learning - CSCL 2025* (pp. 425-429). ISLS.
Ojeda-Ramirez, S., Han, A., Rich, A., Han, S., & Peppler, K. (2025). Reflective Ai-Partnerships: How Middle Schoolers Balance Creativity and Ai Collaboration. In Oshima, J., Chen, B., Vogel, F., & Järvelä, J. (Eds.), *Proceedings of the 18th International Conference on Computer-Supported Collaborative Learning - CSCL 2025* (pp. 404-408). ISLS.
Suthers, D. D., Toth, E., & Weiner, A. (1997). An integrated approach to implementing collaborative inquiry in the classroom. In *the Proceedings of CSCL* - CSCL 1997 (pp. 272-279).
Tabak, I., & Kyza, E. A. (2018) Research on scaffolding in the learning sciences: A methodological perspective. In Fischer, F., Hmelo-Silver, C. E., Goldman, S. R. & Reimann, P. (Eds.), *International Handbook of the Learning Sciences* (pp. 191-200).
Tjønn, H., Ludvigsen, S., & Mørch, A. I. (2025). Generative AI in Norwegian Classrooms: Differences in Teaching About, With and Through Gai. In Oshima, J., Chen, B., Vogel, F., & Järvelä, J. (Eds.), *Proceedings of the 18th International Conference on Computer-Supported Collaborative Learning - CSCL 2025* (pp. 605-607). ISLS.
Wang, R., Zulkifli, N. N., & Mohd Ayub, A. F. (2024). Investigating the Impact of the Stratified Cognitive Apprenticeship Model on High School Students' Math Performance. *Education Sciences, 14*(8), 898.
Warschauer, M., Tseng, W., Yim, S., Webster, T., Jacob, S., Du, Q., & Tate, T. (2023). The affordances and contradictions of AI-generated text for writers of English as a second or foreign language. *Journal of Second Language Writing, 62*.
Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. J*ournal of Child Psychology and Psychiatry, 17*(2), 89–100.
Xin, H., Li, S., Huang, L., Yip, V. W., Niu, Q., Chen, X., & Liu, J. (2025). CocoNote: Agents-Aided Collaborative Learning Environment Enhances Socially Shared Regulation. In Oshima, J., Chen, B., Vogel, F., & Järvelä, J. (Eds.), *Proceedings of the 18th International Conference on Computer-Supported Collaborative Learning - CSCL 2025* (pp. 380-384). ISLS.

Acknowledgements

We sincerely thank the Norwegian Research Council and for the support of the project "Learning in the Age of Algorithms" (#341216) and the Department of Education at the University of Oslo for their financial support.

International Society of the Learning Sciences

Peer Feedback's Role in Undergraduates' Interdisciplinary Collaborative Learning Outcomes and Regulated Learning Patterns

Guo Su, National Institute of Education, Nanyang Technological University, nie22.sg4560@e.ntu.edu.sg
Wenli Chen, National Institute of Education, Nanyang Technological University, wenli.chen@nie.edu.sg
Gaoxia Zhu, National Institute of Education, Nanyang Technological University, gaoxia.zhu@nie.edu.sg
Yubei Shi, National Institute of Education, Nanyang Technological University, nie25.sy@e.ntu.edu.sg
Li Wei, National Institute of Education, Nanyang Technological University, nie25.wl@e.ntu.edu.sg
Chencheng Le, Asian School of the Environment, Nanyang Technological University, ccle@ntu.edu.sg

Abstract: Interdisciplinary collaborative learning (ICL) is critical in 21st-century education, yet supporting disciplinary boundary crossing remains challenging. Peer feedback has the potential to help students cross boundaries by identifying the gaps between single-disciplinary and interdisciplinary approaches. However, the impact of peer feedback on ICL outcomes and regulated learning patterns is underexplored, despite the critical role of regulatory processes in students' engagement in collaborative learning. This study employed a counterbalanced design with 52 undergraduates randomly assigned to two clusters experiencing conditions with and without peer feedback in reverse sequence. Group concept maps and verbal discussions from the 50-minute learning task were collected and analyzed. Results showed that groups performed significantly better with peer feedback in involving diverse perspectives, integrating knowledge, and reflecting on solution impacts and limitations. Additionally, students learning with peer feedback demonstrated an evaluation-engaged regulatory pattern, whereas they showed a planning-engaged regulatory pattern when peer feedback was not given. These findings suggest peer feedback's potential for supporting ICL.

Introduction

Interdisciplinary collaborative learning is a learning process in which individuals bring different sets of skills and knowledge to the teams to achieve a common purpose (Miles & Rainbird, 2015; Su et al., 2026). Introducing ICL into academic and professional curricula helps prepare students to make connections between seemingly fragmented knowledge and to apply that knowledge to real-world problems (Holley, 2017). ICL provides opportunities for students to learn. However, students usually face challenges in utilizing diversity and contradictions of perspectives to cross disciplinary boundaries (Repko, 2008). How to effectively utilize diverse knowledge and perspectives to cross-disciplinary boundaries is a problem to address.

Transformation has been identified as a mechanism that constitutes the learning potential of boundary-crossing (Akkerman & Bakker, 2011). Transformative learning theory suggests that activating disorienting dilemmas, a sense of uncertainty, discomfort, and frustration, can facilitate disciplinary boundary crossings to achieve fresh and integrated understanding (Mezirow, 2000). Research found that disorienting dilemmas can be activated when students encounter interdisciplinary approaches to address complex real-world problems that transcend single-discipline solutions that they are familiar with (Feng et al., 2023). Peer feedback, a communication process through which learners enter dialogues related to performance and standards (Liu & Carless, 2006), among interdisciplinary group members has the potential to activate disorienting dilemmas (Roberts, 2013). Providing and receiving peer feedback helps students identify and reflect discrepancies between peers' and their own current and desired performance levels (Hattie & Timperley, 2007). Existing research on peer feedback in ICL mainly focuses on students' perceived role of peer feedback in the learning process (Wilmot & McKenna, 2018). More empirical research is needed to unpack how peer feedback impacts students' ICL outcomes and processes.

Increasing research has investigated regulated learning mechanisms used in collaborative learning to determine how both groups and individuals initiate, maintain, facilitate, and successfully navigate collaborative learning (Zheng et al., 2019). Two key concepts that describe these collaborative learning processes are self-regulated learning (SRL) and socially shared regulated learning (SSRL) (Hadwin et al., 2011). SRL involves individuals deliberately modifying their cognitive processes, behaviors, motivational states, and emotional responses toward their learning objectives (Zimmerman, 2000). Meanwhile, groups engaged in SSRL use various social processes to regulate their collaborative activities (Volet et al., 2009). However, research has traditionally examined SRL and SSRL separately instead of investigating them jointly (Järvelä et al., 2023), and little is known about how SRL and SSRL co-occur in learning tasks. Furthermore, existing research has primarily concentrated

 © ISLS

on regulatory processes within single-discipline groups, offering minimal understanding of how interdisciplinary groups regulated their learning (O'Donnell et al., 2024). This study aims to explore whether and how peer feedback influences the regulated learning patterns and learning outcomes of undergraduate students engaged in ICL.

Literature review

Interdisciplinary collaborative learning

Shifting from monodisciplinary to interdisciplinary in higher education is motivated by future professional demands and the need to prepare students to deal with complex societal challenges that cannot be addressed by a single discipline (Schijf et al., 2023). Different from interdisciplinary learning, which can be achieved by individuals through integrating knowledge from two or more disciplines, ICL focuses not only on interdisciplinary learning but on collaborative learning in which interdisciplinary work can be associated with collaboration (Verster & van den Berg, 2022). In ICL, team members from different disciplines contribute and integrate their knowledge to study a shared topic collaboratively (Claus & Wiese, 2019). Being exposed to interdisciplinary problems, students not only need to learn about other disciplines but also need to be able to integrate different views and positions.

Existing research on ICL focuses on the enhancement of students' 21st-century skills, academic performance, and knowledge acquisition. Some studies sought to enhance learners' 21st-century skills through ICL. For instance, Tan and Vicente (2019) designed innovative experiential learning by bringing in students from marketing management, computer science, and production design to train their collaboration and teamwork, creativity, critical thinking, and problem-solving skills. In addition, some studies aim to facilitate students' academic performance and knowledge acquisition by designing and implementing ICL. For example, Paredes-Velasco et al. (2023) designed a quasi-experiment to study the impact of interdisciplinary practices and collaboration between Computer Science students and Education students within realistic contexts in university education. The experimental group followed the interdisciplinary realistic project methodology, while the control group followed a more traditional methodology based on collaborative learning. Results supported the effectiveness of the ICL approach in enhancing knowledge acquisition when compared to a more traditional collaborative learning approach. Despite the increasing number of studies on ICL (Heikkinen & Räisänen, 2018), there is still a lack of empirical research on whether pedagogical support will affect students' knowledge integration outcomes and the trajectories of ICL. This gap is particularly concerning given the growing importance of ICL in addressing complex real-world challenges. We need more evidence-based insights into how researchers and educators can effectively foster students' ability to work across disciplinary boundaries and integrate diverse perspectives.

Peer feedback for interdisciplinary collaborative learning

Achieving effective communication and knowledge integration across disciplines is the main challenge of ICL identified in existing literature (Brassler & Dettmers, 2017). Pedagogical support is needed to address the challenges. Working with diverse and unfamiliar areas of expertise, cultures, habits, roles, and knowledge (Edmondson & Harvey, 2018) can be considered a process of crossing boundaries in ICL. Akkerman and Bakker (2011) identified four mechanisms that constitute boundary-crossing, i.e., identification, coordination, reflection, and transformation. Different from the other three mechanisms, which maintain existing boundaries and practices, transformation leads to fundamental changes that can alter or dissolve existing boundaries (Akkerman & Bakker, 2011).

Transformative learning theory suggests that new, integrated conceptual understanding is initiated by disorienting dilemmas (Mezirow, 2000). It happens when students encounter situations in which previous methods of coping or meaning-making become ineffective (Damianakis et al., 2020). During ICL, disorienting dilemmas emerge when students encounter unfamiliar interdisciplinary approaches to address complex, ill-structured real-world problems that transcend single-discipline solutions that they are familiar with (Feng et al., 2023). Based on this, a possible scenario in which peer feedback could activate disorienting dilemmas is when peer assessors evaluate the work of their peers using interdisciplinary criteria, limitations, and weaknesses derived from the gaps between discipline-specific and interdisciplinary approaches. The gap-based peer feedback process serves as a critical mechanism for students to discover gaps between their current performance and expected standards (Boud & Molloy, 2013). When providing feedback, students engage in problem detection, diagnosis, and strategy development (Hayes et al., 1987), which helps them identify discrepancies between current and desired performance levels while offering suggestions for improvement (Hattie & Timperley, 2007; Huisman et al., 2018).

Drawing from existing research, empirical studies specifically examining the role of peer feedback in ICL contexts remain notably limited, despite its potential significance.

Self- and socially shared regulated learning in interdisciplinary collaborative learning

To achieve quality collaborative learning outcomes, groups need to be willing to engage in the co-construction of knowledge and regulate their strategies (Järvelä & Hadwin, 2013). However, previous studies have mainly examined regulated learning in mono-disciplinary groups with limited insights into the ICL contexts (O'Donnell et al., 2024). Homogeneous groups working on interdisciplinary projects (e.g., DiDonato, 2013) differ significantly from interdisciplinary groups. Students taking part in interdisciplinary group work are introduced to new ways of thinking, communicating, and working in a way they would not usually do (Miles & Rainbird, 2015). Furthermore, until relatively recently, research on the regulation of learning has predominantly focused on self-regulated learning (SRL). SRL encompasses metacognitive, cognitive, and behavioral processes to achieve learning goals (Zimmerman, 2015), providing a lens into how individuals regulate their learning process. However, SRL does not adequately address regulation processes within collaborative learning where individuals' learning can be regulated by others (O'Donnell et al., 2024).

When individual students work together, both self- and socially shared regulated learning contribute to productive collaboration (Hadwin et al., 2018). SRL happens when students actively manage their own learning using cognitive, metacognitive, and motivational strategies (Zimmerman, 2013). These strategies are used to set personal learning goals, monitor progress, implement effective learning approaches, and reflect on outcomes to improve future learning. Zimmerman (2013) characterized self-regulated learners as proactive individuals who effectively monitor and skillfully adjust their behaviors to achieve positive learning outcomes. SSRL goes beyond the individual to group settings, focusing on how group members collaboratively regulate their learning processes. In SSRL, group members work together to establish shared goals, monitor collaborative understanding, and change strategies to achieve learning outcomes effectively (Michalsky & Cohen, 2021). These collaborative regulatory processes enable teams to coordinate their interactions, overcome challenges, and maintain productive collaborative engagement. Research has traditionally examined individual regulatory processes (SRL) and group regulatory processes (SSRL) separately instead of investigating them jointly. For example, Lobczowski et al. (2020) investigated SRL patterns in collaborative learning contexts and how individual regulation results in learning outcomes.

The joint impact of SRL and SSRL on group performance, specifically within collaborative learning contexts, is still underexplored (Zheng et al., 2019). Understanding the co-occurrence of SRL and SSRL behaviors in ICL is crucial because the dynamic interaction between individual and shared regulation significantly influences learners' ability to collaborate effectively. Recent research has started to address these challenges through innovative methodologies and analytical methods. Ordered Network Analysis (ONA) emerges as a particularly promising analysis method for investigating the temporal dynamics of SRL and SSRL in collaborative argumentation. ONA enables researchers to map and analyze sequential patterns and relationships within learning interactions, providing insight into how individual and collective regulation processes unfold over time (Fan et al., 2023). In the context of collaborative learning, ONA visualizes and quantifies how learners interact, share resources, resolve conflicts, and regulate their learning processes both individually and collectively (Nguyen et al., 2022). This approach is especially valuable for capturing the dynamic interplay between self-regulation and socially shared regulation in collaborative environments.

Taken together, this study aims to investigate whether and how peer feedback impacts undergraduate students' regulated learning patterns and learning outcomes in ICL. Specifically, the research questions are: (1) What are the differences in the quality of the final products between groups learning with and without peer feedback in interdisciplinary collaborative learning? (2) How do regulated learning patterns differ between undergraduate students learning with and without peer feedback?

Methods

Participants and context

A counterbalanced design was implemented to investigate the effects of peer feedback on undergraduate students' quality of interdisciplinary work and regulated learning patterns in ICL. In this study, the participatory groups were randomly assigned to Cluster 1 with 5 groups and Cluster 2 with 4 groups. Groups in cluster 1 experienced the experiment condition (i.e., with peer feedback) before the control condition (i.e., without peer feedback).

A total of 52 undergraduate students (29 males and 23 females) from different disciplinary backgrounds, such as computer science, engineering, business, and psychology, participated in this research. All participants participated in the research based on their own willingness. Most of them (N=42) were second-year students, and

some were in their fourth year (N=7), third year (N=2), and first year (N=1). They enrolled in a compulsory ICL course at a university in Singapore. Students were randomly assigned to nine interdisciplinary groups of 5 or 6. In this course, students were required to analyze current real-world challenges regarding sustainable development goals (SDGs), such as "how can we coexist with macaques in Singapore to create a sustainable city?". To ensure the comparability of the difficulty, the topics were all about SDGs and were selected by the course instructor, with four years of experience teaching this course.

This study took place in a multimedia-enabled tutorial classroom, where group members gathered around round tables equipped with electronic display screens. Students utilized their personal mobile devices while engaging in face-to-face discussions and simultaneously creating concept maps to visualize their group solutions through Miro (https://miro.com/). Miro is a collaborative learning platform, providing teams with a dynamic space to develop visual representations and foster collaborative ideation and problem-solving. The platform's real-time collaboration features, including synchronous editing, enable seamless interaction among team members.

Research design

Figure 1 shows the learning task design throughout the study sessions. First, individuals spent 15 minutes creating their own concept maps based on their disciplinary background knowledge. After that, students under the control condition took turns to share with other group members their ideas on the individual concept maps. By contrast, students under the experimental condition were required to give feedback to another group member from a different discipline. The feedback providers and receivers were randomly assigned by the researchers. Based on a peer assessment model developed by Sluijsmans and Van Merrie¨nboer (2000), students need to (1) define assessment criteria—think about what is required; (2) judge the quality of a peer's product; and (3) provide feedback for future learning—give constructive feedback about the work produced by a peer. The assessment criteria were provided to guide assessors to identify gaps between their peers' current discipline-bound approaches and the desired interdisciplinary performance that requires integrating multiple perspectives (see Table 1). Adapted from Gan and Hattie's (2014) guiding questions for peer feedback generation, students needed to answer the three questions to give peer feedback in 15 minutes: (1) What are the expectations of an interdisciplinary approach to address the problem, and what are the differences between a disciplinary approach and an interdisciplinary approach to solve the problem? (2) What didn't he/she do well, according to the interdisciplinary criteria and differences identified? Give explanations to support your feedback, and (3) How can he/she improve the current piece of work, based on the interdisciplinary criteria? Give explanations to support your feedback.

Figure 1

The Learning Task Design

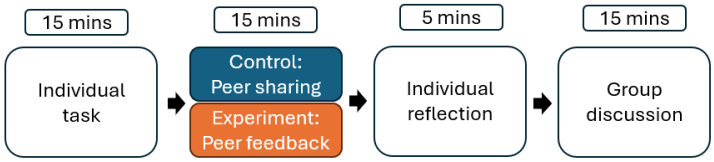


Table 1

The Assessment Criteria of Individual Concept Maps for Peer Feedback

	Novice level	Intermediate level	Mastery level
Problem framing	The map does not contain an identifiable purpose, or the purpose is unclear.	The student clearly states a purpose that calls for an integrative approach.	The student clearly states a purpose that calls for an integrative approach and provides a clear rationale for taking this interdisciplinary approach.
Number of integrations	No evidence of disciplinary integration (e.g., no links between disciplines).	Students raised evidence of disciplinary integration.	Students raised two or more pieces of evidence of disciplinary integration.
Depth of integration	No explanations for how the disciplines selected support/intertwine with each other.	List disciplinary concepts only, without building off the concepts with each other.	Leverage contributing disciplinary concepts by building off each other to effectively address the problem/task in a way that one discipline cannot.

 © ISLS

After peer sharing or peer feedback, students under all the conditions individually reflected on the limitations of their perspectives and assumptions, how they are aware of the limitations, and why and how they should address the limitations to make their assumptions and perspectives more sustainable. Finally, group members discuss together and generate a group concept map by integrating ideas from different disciplines. To control the extraneous variables, the instructor for the two clusters was the same, and the time duration of each step in each condition was kept the same.

Data collection and analysis methods

Each group's generated concept maps (18 concept maps generated by nine groups in the two sessions) were gathered as the final products of their ICL. In addition, group audio discussion data during each session was collected. To answer the first research question, we adopted a rubric for assessing the quality of group reports adapted from Mansilla et al.'s (2009) criteria and Tripp and Shortlidge's (2020) rubric (Su et al., 2024; 2025) (https://drive.google.com/file/d/1ggzl_BWq5O5c2N6eJgSXZZF8lHic77sC/view?usp=sharing).

The four dimensions of the rubric were purposefulness, disciplinary grounding, advancement through integration, and critical awareness, and there were two sub-dimensions under each dimension. Under purposefulness, students are assessed on whether they frame problems with an integrative approach and their level of stakeholder consideration. The disciplinary grounding dimension examines the range of disciplinary perspectives students incorporate and their ability to explain how different disciplines contribute to solutions. Advancement through integration dimension includes the number of disciplines integrated and the depth of that integration, evaluating whether students merely list contributions or actively leverage and build upon different disciplinary methods to address problems effectively. Finally, critical awareness assesses whether students identify local and broader societal effects and can articulate potential barriers with resolutions. Two authors scored 4 concept maps (22% of the 18 concept maps) individually. The inter-rater reliability (Cronbach's Alpha) of the four dimensions was .928, .882, .951, and .985, respectively. The two raters discussed the differences in scoring. The first author evaluated the remaining 14 maps. After assessing the quality of reports, a Shapiro-Wilk test was conducted to explore the normality of the data. Considering the data was normally distributed ($p = .192 > .05$), a paired-sample t-test was used to compare the differences in the quality of the ICL product between the two conditions.

To answer the second research question, the verbal discussion of each group was first segmented into utterances as the unit of analysis. Each utterance was coded using a coding scheme adapted from Zheng et al. (2019) and van der Graaf et al. (2022) (see Table 2). The coding framework consists of Orientation, Planning, Monitoring, and Evaluation elements at both individual and group levels. A total of 764 utterances were coded. Two coders coded 145 utterances (about 19% of the dataset) independently, and the inter-coder reliability was good (Cohen's Kappa = .855). The remaining 619 utterances were coded by one of the two coders. Then, an ONA (www.epistemicnetwork.org, Tan et al., 2022) was conducted to model and visualize the temporal co-occurrence of the SRL and SSRL patterns demonstrated by students under the two conditions. The stanza size was four, which means that the co-occurrence of collaborative learning codes in every 4 utterances. Previous research found that the stanza size of four is an appropriate value for synchronous discourse-natured data (Ruis et al., 2019).

Table 2

The Coding Scheme for SRL and SSRL

Dimension	Description
SRL - Orientation (SO)	Orientation on the individual learning tasks or on prior knowledge; feelings about the task
SRL - Planning (SP)	Planning of the individual learning process by arranging activities and determining strategies; proceeding to the next topic
SRL - Monitoring (SM)	Monitoring and checking one's own learning process; checking individual progress according to the instruction or plan
SRL - Evaluating (SE)	Checking content-wise correctness of individual learning activities; reflecting on individual learning
SSRL - Orientation (SSO)	Orientation on the group task or on the collaborative learning settings
SSRL - Planning (SSP)	Planning of other group members' learning process by arranging activities and determining strategies

 © ISLS

SSRL - Monitoring (SSM)	Monitoring and checking the group's learning process; solving conflicts and achieving mutual understanding
SSRL - Evaluating (SSE)	Checking content-wise correctness of group learning activities; agreeing or disagreeing with others' ideas; reflecting on group learning

Results

RQ1: Comparison of the quality of group concept maps between the two conditions

To compare the quality of group ICL products, paired-sample t-tests were conducted, with results presented in Table 3. Significant differences existed between the two conditions in the overall score ($t = -3.268$, $p < .05$). Specifically, groups in the experimental condition (learning with peer feedback) significantly outperformed the same groups in control condition (learning without peer feedback) across three dimensions: disciplinary grounding ($t = -4.364$, $p < .01$), integration ($t = -3.317$, $p < .05$), and critical awareness ($t = -3.847$, $p < .01$) with large effect sizes.

These findings indicate that peer feedback enhanced groups' consideration of diverse disciplinary knowledge and perspectives, as well as their ability to reason about the involvement of this knowledge and these perspectives, compared to learning without peer feedback. Furthermore, groups in the experimental condition showed more evidence of knowledge integration and demonstrated more in-depth knowledge integration than when peer feedback was not provided. Additionally, groups demonstrated deeper analysis and reflection on the social impacts and limitations of their proposed solutions after engaging in peer feedback compared to peer sharing. However, groups in the control condition showed slightly better performance in framing problems using integrative approaches and identifying stakeholders, though this difference was not statistically significant.

Table 3

The Result of Paired-Sample T-Tests

	Control Condition		Experiment Condition		t	p	Cohen's d
	Mean	SD	Mean	SD			
Total score	39.000	10.283	50.556	5.503	-3.268	.011*	-1.089
Purposefulness	12.222	3.930	11.333	1.323	.849	.420	.283
Disciplinary Grounding	10.889	2.261	14.222	.972	-4.364	.002**	-1.455
Integration	12.000	3.428	15.667	1.658	-3.317	.011*	-1.106
Critical Awareness	3.889	2.619	9.333	3.162	-3.847	.005**	-1.282

Note. *p < .05, **p < .01

RQ2: Comparison of regulation patterns in interdisciplinary collaborative learning between two conditions

Figure 2 illustrates the comparisons of ONA networks between the two conditions. The red lines indicate a stronger co-occurrence of SRL and SSRL behaviors under the control condition than under the experimental condition. By contrast, the blue lines represent stronger connections between the regulatory behaviors under the experimental condition than under the control condition. The red rectangle represents the centroids (mean position of the projected points) of the groups learning without peer feedback, whereas the blue rectangle represents the centroids of the groups learning with peer feedback. Along the X axis (MR1), a Mann-Whitney test showed that the control condition (Mdn = 0.18, N = 9) was statistically significantly different at the alpha = 0.05 level from the experimental condition (Mdn = -0.14, N = 9, U = 68.00, p = 0.01, r = -0.68) in terms of the distribution of projection points.

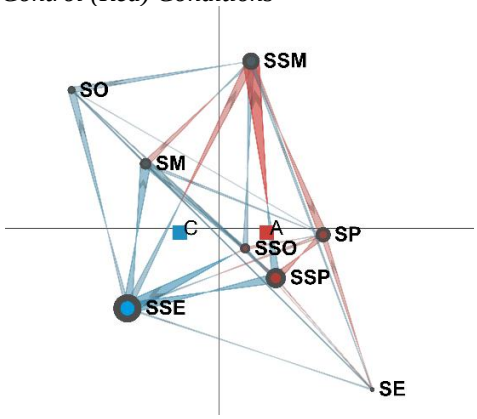
The colored circles within each node represent self-loops when a group continues to stay in an SRL or SSRL activity. The color and saturation of the circle within each node are proportional to the number of self-connections for that code. In the comparison diagram, colored circles that are larger and more saturated reflect codes with more frequent self-loops (Tan et al., 2022). The larger and more saturated blue color of a node indicates a higher frequency of self-loops of the corresponding behavior in the experimental condition, vice versa. According to Figure 2, groups in the experimental condition engaged in more self-loops of SSE and SSM. Meanwhile, groups in the control condition engage in more SSP and SP. The line between codes represents the co-occurrence of regulatory behaviors. Regarding the chevron, for example, between 2 codes SSM and SSP, the thick and saturated triangle with a chevron on it represents the unit's response with code SSM to code SSP. Between any pair of codes, if there is a bidirectional connection, the chevron only appears on the side with stronger

connections, which helps viewers differentiate heavier edges (Tan et al., 2022). Each edge is color-coded to indicate which of the two networks contains the stronger connection. As the thick red lines show, group students under the control condition tended to make plans for themselves (SP) and other team members (SSP) when monitoring the learning process in the group monitoring (SSM). These patterns suggest that control groups followed a more traditional regulatory pattern, where monitoring activities prompted planning responses at both the individual and group levels. It reflects that learners adjusted their strategies primarily in response to identified needs through monitoring. By contrast, groups in the experimental condition exhibited more complex regulatory patterns, as indicated by the prominent blue lines. These groups demonstrated strong co-occurrence among SSE, SSP, SSO, and SM, suggesting a more integrated regulatory process. Students in the experimental condition were more likely to engage in evaluative reflection on their learning goals and progress across both individual and collective dimensions, indicating a proactive rather than reactive approach to regulation.

Notably, both conditions exhibited frequent transitions between SSE and SSM, and between SSP and SSM, highlighting the central role of monitoring in ICL. However, the direction of these transitions differed between conditions. Groups in the experimental condition demonstrated a strong transition from SSM to SSE, suggesting they used monitoring insights to activate reflection. Conversely, groups in the control condition exhibited the reverse pattern (SSE → SSM), moving from evaluation back to monitoring. Similarly, experimental groups showed a transition from SSM to SSP, while control groups followed the opposite sequence (SSP → SSM). These contrasting patterns indicated that experimental groups moved from monitoring to higher-order cognitive processes (evaluating and planning), whereas control groups tended to use evaluation and planning as precursors to monitoring, reflecting fundamentally different regulatory patterns.

Figure 2

The Comparison of Regulation Patterns between Experiment (Blue) and Control (Red) Conditions



Discussion

This study aimed to investigate whether and how peer feedback impacts undergraduate students' ICL outcomes and regulation patterns in the learning process. Specifically, the groups' generated concept maps and group verbal discussions were analysed. The results showed that groups performed significantly better when they learned with peer feedback than without. Moreover, the ONA results illustrated that groups under the experiment condition (with peer feedback) demonstrated more self-loop SSRL-Evaluating and SSRL-Monitoring behaviours than under the control condition. In contrast, groups learning without peer feedback showed more SSRL-Planning and SRL-Planning behaviours.

Regarding the comparison of ICL outcomes between the two conditions, groups learning with peer feedback outperformed those groups learning without peer feedback in disciplinary grounding, knowledge integration, and critical awareness. These results can be attributed to peer feedback in facilitating the incorporation of diverse perspectives supported by explicit reasoning. When students engaged in peer feedback, they became aware of gaps in their own understanding and examined their peers' approaches (Chen et al., 2023; Tan et al., 2023). This awareness prompted students to develop more detailed explanations and justifications to address identified shortcomings and deepen their understanding (Chen et al., 2024). Consequently, groups with peer feedback demonstrated evidence-based reasoning and deep knowledge integration after knowing that their previous solutions based on a single discipline were insufficient to address a complex problem. Furthermore, peer feedback enabled groups to critically reflect on the social impact and limitations of their collaborative work. This reflective capacity likely stems from the nature of peer feedback as a collaborative activity between individuals,

which functions as reflective engagement influencing both feedback givers and receivers (Chen et al., 2024). Such reciprocal reflection promoted deeper critical thinking about the strengths and limitations of proposed interdisciplinary solutions. Nevertheless, no significant difference was found in purposefulness, i.e., indicating and justifying the purpose of using interdisciplinary approaches to address the real-world problems between two conditions. Possibly, this finding reflected the complex nature of purposefulness, which challenges students to present their object of study in ways that are both viable in scope and clearly multidimensional (Mansilla et al., 2009). Students may need more instructions or prompts, beyond what a single cycle of peer feedback can provide.

Additionally, the network analysis revealed distinct regulatory patterns between experimental and control conditions. Groups receiving peer feedback demonstrated more frequent self-loops in SSE and SSM than groups learning without peer feedback and exhibited transitions from monitoring to evaluation (SSM → SSE). The emphasis on self-transitions within SSE suggests that students are repeatedly engaged in evaluation processes, which is consistent with research by Panadero et al. (2019), who found that feedback interventions enhance students' evaluative judgment capabilities. Furthermore, the sequential pattern (SSM → SSE) demonstrated by the experimental condition aligns with Zimmerman's (2000) cyclical model of regulated learning. The model positions that reflection occurs after efforts to learn and is intended to optimize a person's reactions to his or her outcomes. In contrast, control groups exhibited more frequent self-loops in planning behaviors and demonstrated transitions from evaluation and planning back to monitoring (SSE → SSM), reflecting what Zimmerman (2013) describes as a reactive regulatory approach. These reactive groups rely on post-performance self-reflections to learn. But this post hoc focus is hypothesized to diminish their learners' effectiveness.

Limitations and implications

This study has several limitations. First, while the study examines the impact of peer feedback on ICL outcomes and regulation patterns, it does not investigate how the quality and nature of peer feedback influence these outcomes and patterns. Future studies can unpack whether and how variations in peer feedback quality and nature shape ICL. Second, the data sources are limited to verbal discussion and learning outcomes. Future research may include additional data sources and analysis methods to achieve a more comprehensive understanding of the learning process, such as adding emotional and behavioral perspectives through physiological sensors, video analysis, or self-report measures. Third, the relatively small sample size limits generalizability. Future research can include larger participant samples and employ longitudinal designs to investigate the sustained effects of peer feedback on ICL over time.

This study implies that designing and implementing peer feedback among interdisciplinary group members can facilitate involving more disciplinary knowledge with strong knowledge integration and reflecting on the impact and limitations of interdisciplinary group work. Additionally, learning with peer feedback promotes proactive evaluation-oriented regulated learning. Educators and researchers can support feedback between students from different disciplines to enhance interdisciplinary task performance and regulated learning in ICL contexts.

References

Akkerman, S. F., & Bakker, A. (2011). Boundary Crossing and Boundary Objects. *Review of Educational Research*, *81*(2), 132–169.

Boud, D., & Molloy, E. (2013). Rethinking models of feedback for learning: the challenge of design. *Assessment & Evaluation in higher education*, *38*(6), 698-712.

Brassler, M., & Dettmers, J. (2017). How to enhance interdisciplinary competence—interdisciplinary problem-based learning versus interdisciplinary project-based learning. *Interdisciplinary Journal of Problem-Based Learning*, *11*(2).

Chen, W., Hu, H., Lyu, Q., & Zheng, L. (2024). Using peer feedback to improve critical thinking in computer-supported collaborative argumentation: An exploratory study. *Journal of Computer Assisted Learning*, *40*(6), 3390-3405.

Chen, W., Lyu, Q., Su, J., Aileen, C. S. C., Zhang, W., Su, G., & Li, X. (2023). How does feedback formulation pattern differ between more-improvement and no-improvement student groups? An exploratory study. In *Proceedings of the 17th International Conference of the Learning Sciences-ICLS 2023, pp. 577-584*. International Society of the Learning Sciences.

Claus, A. M., & Wiese, B. S. (2019). Development and test of a model of interdisciplinary competencies. *European Journal of Work and Organizational Psychology*, *28*(2), 191-205.

Damianakis, T., Barrett, B., Archer-Kuhn, B., Samson, P. L., Matin, S., & Ahern, C. (2020). Transformative learning in graduate education: Masters of social work students' experiences of personal and professional learning. *Studies in Higher Education*, *45*(9), 2011–2029.

DiDonato, N. C. (2013). Effective self-and co-regulation in collaborative learning groups: An analysis of how students regulate problem solving of authentic interdisciplinary tasks. *Instructional science*, *41*(1), 25-47.

Edmondson, A. C., & Harvey, J. F. (2018). Cross-boundary teaming for innovation: Integrating research on teams and knowledge in organizations. *Human Resource Management Review*, *28*(4), 347-360.

Fan, Y., Tan, Y., Raković, M., Wang, Y., Cai, Z., Shaffer, D. W., & Gašević, D. (2023). Dissecting learning tactics in MOOC using ordered network analysis. *Journal of Computer Assisted Learning*, 39(1), 154-166.

Feng, X., Sundman, J., Aarnio, H., Taka, M., Keskinen, M., & Varis, O. (2023). *To Withdraw, Investigate, Negotiate Or Integrate? Students' Coping Strategies With Disorienting Dilemmas In Interdisciplinary Project Courses*.

Gan, M. J. S., & Hattie, J. (2014). Prompting secondary students' use of criteria, feedback specificity and feedback levels during an investigative task. *Instructional Science*, *42*(6), 861–878.

Hadwin, A. F., Järvelä, S., & Miller, M. (2011). Self-regulated, co-regulated, and socially shared regulation of learning. *Handbook of Self-regulation of Learning and Performance*, *30*, 65-84.

Hadwin, A. F., Järvelä, S., & Miller, M. (2018). Self-regulation, co-regulation, and shared regulation in collaborative learning environments: A social cognitive perspective. In D. H. Schunk, & J. A. Greene (Eds.), *Handbook of self-regulation of learning and performance* (pp. 83–106). Routledge.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112.

Hayes, S. C., Nelson, R. O., & Jarrett, R. B. (1987). The treatment utility of assessment: A functional approach to evaluating assessment quality. *American Psychologist*, *42*(11), 963.

Heikkinen, K. P., & Räisänen, T. (2018). Role of multidisciplinary and interdisciplinary education in computer science: A literature review.

Holley, K. (2017). Administering interdisciplinary programs. In R. Frodeman, J. T. Klein, & R. Pacheco (Eds.), *The Oxford handbook of interdisciplinarity* (pp. 530–543). Oxford University Press.

Huisman, B., Saab, N., Van Driel, J., & Van Den Broek, P. (2018). Peer feedback on academic writing: undergraduate students' peer feedback role, peer feedback perceptions and essay performance. *Assessment & Evaluation in Higher Education*, *43*(6), 955-968.

Järvelä, S., & Hadwin, A. F. (2013). New frontiers: Regulating learning in CSCL. *Educational Psychologist*, *48*(1), 25–39.

Järvelä, S., Nguyen, A., Vuorenmaa, E., Malmberg, J., & Järvenoja, H. (2023). Predicting regulatory activities for socially shared regulation to optimize collaborative learning. *Computers in Human Behavior*, *144*, 107737.

Liu, N. F., & Carless, D. (2006). Peer feedback: the learning element of peer assessment. *Teaching in Higher education*, *11*(3), 279-290.

Lobczowski, N. G., Allen, E. M., Firetto, C. M., Greene, J. A., & Murphy, P. K. (2020). An exploration of social regulation of learning during scientific argumentation discourse. *Contemporary Educational Psychology*, *63*, 101925.

Mansilla, V. B., Duraisingh, E. D., Wolfe, C. R., & Haynes, C. (2009). Targeted assessment rubric: An empirically grounded rubric for interdisciplinary writing. *The Journal of Higher Education*, *80*(3), 334-353.

Mezirow, J. (2000). Learning to think like an adult. In J. Mezirow & Associates (Eds.), *Learning as transformation: Critical perspectives on a theory in progress*. Jossey-Bass.

Michalsky, T., & Cohen, A. (2021). Prompting socially shared regulation of learning and creativity in solving STEM problems. *Frontiers in Psychology*, *12*.

Miles, M., & Rainbird, S. (2015). Evaluating interdisciplinary collaborative learning and assessment in the creative arts and humanities. *Arts and Humanities in Higher Education*, *14*(4), 409-425.

Nguyen, A., Järvelä, S., Wang, Y., & Rosé, C. P. (2022). Exploring socially shared regulation with an AI deep learning approach using multimodal data. In Chinn, C., Tan, E., Chan, C., & Kali, Y. (Eds.), *Proceedings of the 16th International Conference of the Learning Sciences - ICLS 2022* (pp. 527-534). International Society of the Learning Sciences.

O'Connell, M. T., Stöhr, C., Wallin, P., & Negretti, R. (2024). Social regulation of learning in interdisciplinary groupwork. *European Journal of Engineering Education*, *49*(4), 683-699.

Panadero, E., Broadbent, J., Boud, D., & Lodge, J. M. (2019). Using formative assessment to influence self-and co-regulated learning: the role of evaluative judgement. *European Journal of Psychology of Education*, *34*(3), 535-557.

Paredes-Velasco, M., Arnal-Palacián, M., Urquiza-Fuentes, J., & Martín-Lope, M. (2023). Improving Soft Skills Through an Interdisciplinary Approach in a Realistic Context Between Education and CS Students in an HCI Course. *IEEE Transactions on Education*.

Repko, A. F. (2008). *Interdisciplinary research*. Thousand Oaks, CA: SAGE.

Roberts, N. (2013). Disorienting dilemmas: Their effects on learners, impact on performance, and implications for adult educators.

Ruis, A., Siebert-Evenstone, A., Pozen, R., Eagan, B., & Shaffer, D. W. (2019). Finding common ground: A method for measuring recent temporal context in analyses of complex, collaborative thinking.

Schijf, J. E., van der Werf, G. P., & Jansen, E. P. (2023). Measuring interdisciplinary understanding in higher education. *European Journal of Higher Education*, *13*(4), 429-447.

Sluijsmans, D. M. A., Brand-Gruwel, S., Van Merrie¨nboer, J. J. G., & Bastiaens, T. J. (2002). The training of peer assessment skills to promote the development of reflection skills in teacher education. *Studies in Educational Evaluation*, *29*(1), 23–42.

Su, G., Chen, W., Zhu, G., & Le, C. (2026). Scripting Undergraduates' Interdisciplinary Collaborative Learning to Enhance Their Competence. In *Handbook of Asian Educational Innovation towards the Futures of Education* (pp. 1-18). Singapore: Springer Nature Singapore.

Su, G., Chen, W., Zhu, G., Le, C., & Zheng, L. (2024). Scripting Undergraduates' Interdisciplinary Collaborative Learning to Enhance Their Interdisciplinary Competence. In *Proceedings of the 17th International Conference on Computer-Supported Collaborative Learning-CSCL 2024, pp. 265-268*. International Society of the Learning Sciences.

Su, G., Chen, W., Zhu, G., Xu, J. S., & Le, C. (2025). Exploring Students' Interdisciplinary Collaborative Learning Outcomes and Processes with Varying Degrees of Knowledge Heterogeneity. In *Proceedings of the 18th International Conference on Computer-Supported Collaborative Learning-CSCL 2025, pp. 326-329*. International Society of the Learning Sciences.

Tan, J. S., Chen, W., Su, J., & Su, G. (2023). The mechanism and effect of class-wide peer feedback on conceptual knowledge improvement: Does different feedback type matter?. *International Journal of Computer-Supported Collaborative Learning*, *18*(3), 393-424.

Tan, T. A. G., & Vicente, A. J. (2019). An innovative experiential and collaborative learning approach to an undergraduate marketing management course: A case of the Philippines. *The International Journal of Management Education*, *17*(3), 100309.

Tan, Y., Ruis, A. R., Marquart, C., Cai, Z., Knowles, M. A., & Shaffer, D. W. (2022, October). Ordered network analysis. In *International Conference on Quantitative Ethnography* (pp. 101-116). Cham: Springer Nature Switzerland.

Tripp, B., & Shortlidge, E. E. (2019). A framework to guide undergraduate education in interdisciplinary science. *CBE—Life Sciences Education*, *18*(2), es3.

Van der Graaf, J., Lim, L., Fan, Y., Kilgour, J., Moore, J., Gašević, D., ... & Molenaar, I. (2022). The dynamics between self-regulated learning and learning outcomes: An exploratory approach and implications. *Metacognition and Learning*, *17*(3), 745-771.

Verster, B., & van den Berg, C. (2022). Theorising with sociomateriality: Interdisciplinary collaboration in socio-technical learning environments. *Educational Research for Social Change*, *11*(2), 39-56.

Wilmot, K., & McKenna, S. (2018). Writing groups as transformative spaces. *Higher Education Research & Development*, *37*(4), 868-882.

Zheng, J., Xing, W., & Zhu, G. (2019). Examining sequential patterns of self-and socially shared regulation of STEM learning in a CSCL environment. *Computers & Education*, *136*, 34-48.

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. Pintrich, & M. Zeidner (Eds.), Handbook of self-regulation (pp. 13–39). San Diego, CA: Academic Press.

Zimmerman, B. J. (2013). From cognitive modeling to self-regulation: A social cognitive career path. *Educational psychologist*, *48*(3), 135-147.

Zimmerman, B. J. (2015). Self-regulated learning: Theories, measures, and outcomes. *International encyclopedia of the social & behavioral sciences*, 541-546.

Acknowledgements

This study was funded by Nanyang Technological University (NTU), Singapore (EdeX Teaching and Learning grant #023339-00001) and Singapore Ministry of Education (MOE) Tertiary Education Research Grant (MOE2024-TRF-008), both administered by NTU. Any opinions, conclusions or implications in this material are those of the author(s) and do not necessarily reflect the views of NTU. The study was approved by the NTU IRB (number: IRB-2023-048).

International Society of the Learning Sciences

How Do Pre-Service Teachers Integrate Multiple Peer Feedback Messages? Identifying Different Integration Patterns

Julia Hornstein, Martin Greisel, and Ingo Kollar
julia.hornstein@uni-a.de, martin.greisel@uni-a.de, ingo.kollar@uni-a.de
University of Augsburg, Germany

Abstract: Computer-supported peer feedback is a valuable approach to enhancing student learning in CSCL environments. Yet, little is known about how learners process multiple peer feedback messages. In an exploratory study with $N = 100$ pre-service teachers who analyzed classroom cases, provided feedback, and integrated feedback from two peers, we aimed to identify different feedback integration patterns (RQ1) and investigated how these are predicted by objective feedback features and perceived feedback quality (RQ2). A hierarchical cluster analysis yielded three integration patterns: a confirmation pattern, a conflict pattern, and a selective integration pattern. Multinomial logistic regressions revealed that objective feedback features—specifically accuracy, relevance, and implementability—predicted these integration patterns, whereas perceived feedback quality did not. These findings highlight that integrating multiple peer feedback messages is a complex, feature-dependent process. In addition, they offer implications for designing adaptive, computer-supported peer feedback systems in CSCL settings.

Aims of the study

Peer feedback, conceptualized as a collaborative learning process, has been extensively examined within the area of Computer-Supported Collaborative Learning (CSCL) and has consistently been demonstrated to facilitate student learning (Özbek et al., 2024; Zong et al., 2021). Although current research provides strong evidence that providing feedback constitutes a powerful mechanism for learning, how learners actually use the feedback they receive from their peers effectively remains rather unclear. This is especially the case when learners receive feedback from more than one peer, that is when they need to integrate multiple peer feedback messages (Lundstrom & Baker, 2009; Nicol et al., 2014). Nevertheless, systematic theoretical frameworks and empirical investigations elucidating how learners integrate multiple peer feedback messages remain scarce.

Drawing on research concerning learning from multiple documents (List & Alexander, 2019), we posit that the optimal integration of multiple peer feedback messages entails four steps: identification, separate representation, simultaneous relation, and relational designation. After relationships between the multiple peer feedback messages have been identified, students need to connect these multiple peer feedback messages through synthesis and reconciliation (List & Alexander, 2019). Yet, it remains unclear (a) to what extent and in what ways learners engage in these integration processes when integrating multiple peer feedback messages, and (b) how both objective feedback features (relevance, accuracy, implementability, structural quality) and the perceived quality of the received feedback (adequacy, subjective quality) influence this integration process. The present study seeks to provide empirical evidence to address these questions.

Peer feedback in the context of CSCL

The peer feedback process basically consists of four phases (Kollar & Fischer, 2010). In the *task performance* phase, students work on a task and produce an initial task solution (e.g., a written case analysis). These initial task solutions represent the product on which peers then formulate feedback in the *feedback provision* phase. This is followed by the *feedback reception* phase, during which students process the feedback they have received and compare it to their initial task solutions. The feedback process ends with the *revision* phase, in which students revise their initial task solution based on the feedback they have received.

In the context of CSCL, peer feedback is regarded as an important instructional method, as learners collaborate and digital technologies can support the peer feedback process (Zong et al., 2021). From a sociocultural perspective (Vygotsky et al., 1978), students in the peer feedback process may provide support to each other's Zone of Proximal Development (ZPD), enabling them to operate in their ZPD and thus advance their actual development. This is achieved by helping each other through asking questions and offering suggestions for possible revisions. From a socio-cognitive perspective, the reason for the effectiveness of peer feedback lies in the fact that important cognitive processes take place during both feedback provision and feedback reception (Nicol & McCallum, 2022; van Popta et al., 2017).

 © ISLS

Compared to analogue forms of peer feedback, computer-supported peer feedback offers several advantages, such as greater flexibility in application (both synchronous and asynchronous) and the ability to randomly, anonymously, and automatically assign feedback pairs (Chen, 2016; Cho & Schunn, 2007; Panadero & Alqassab, 2019). Furthermore, computer-supported peer feedback processes can easily be structured by incorporating instructional support into each phase (e.g., through the provision of assessment criteria, rubrics, or training; Alqassab et al., 2018; Dmoshinskaia et al., 2021) and easily monitored by the teacher (Farahani et al., 2019). Several empirical studies have confirmed these advantages in CSCL contexts, showing that computer-supported peer feedback processes can positively impact student performance (Özbek et al., 2024; Zong et al., 2021), particularly when supported by instruction (Hornstein, Keller et al., 2025).

Integrating multiple peer feedback messages

Although the feedback reception phase in the peer feedback process is conceptually complex and has received relatively limited empirical attention, it offers considerable potential for promoting student learning (Nicol et al., 2014). When learners engage with feedback from their peers, they need to compare their initial task solutions with the feedback messages received. This comparison exposes them to alternative perspectives, encourages critical reflection on their prior understanding, and potentially supports the refinement of their initial task solution (Nicol & McCallum, 2022). Such learning benefits may even be larger when students receive feedback messages from multiple peers rather than from a single peer, as integrating multiple peer feedback messages is likely to involve even deeper cognitive processing (Cho & Schunn, 2007). However, despite their importance, the nature of these integration processes remains insufficiently explored.

Empirical investigations into how students process and learn from multiple peer feedback messages are still scarce. In one study, Li et al. (2010) provided each participant with feedback from two peers, explicitly noting that feedback quality might vary and that conflicting suggestions could occur. Students were therefore required to compare both feedback messages and decide which comments to incorporate into their revisions. Interestingly, the study found no significant relationship between the quality of the feedback participants received and subsequent revision quality. By contrast, Cho and Schunn (2007) reported that students who received feedback from multiple peers produced essays of higher quality than those who received feedback from only one peer. More recently, van Meenen et al. (2023) examined how students process discrepant peer feedback using eye-tracking data and simulated feedback comments. Their results revealed pronounced individual differences: while some learners recognized inconsistencies but disregarded them, others actively attempted to reconcile these discrepancies. These findings indicate that students do not automatically engage effectively with multiple peer feedback messages; instead, they display diverse integration patterns, ranging from largely ignoring the feedback to actively resolving conflicting information. However, despite these initial findings, it is still unclear exactly how (high-quality) integration of multiple peer feedback messages looks like.

To advance our understanding of how learners integrate feedback from multiple peers, we used the Integrated Framework of Multiple Text Use (List & Alexander, 2019). This framework describes how individuals engage with, process, and construct meaning from multiple texts and comprises three stages: preparation, execution, and production. In the *preparation* stage, learners orient themselves toward the task by considering external task features, personal perceptions, and individual factors such as prior knowledge, which shape their cognitive engagement. During the *execution* stage, students employ metacognitive monitoring and cognitive strategies—including comprehension and cross-textual linking—to connect and integrate information across texts until they determine whether their cognitive goals are met. Finally, in the *production* stage, learners translate the cognitive and affective outcomes of text processing into external products, such as written work, which may or may not fully reflect their internal conceptual, emotional, or belief-based changes.

Within this model, processing activities—particularly those related to information integration—play a central role. A key process identified by List and Alexander (2019) is cross-textual linking, which involves establishing meaningful connections between multiple texts to ensure mindful integration (Berndt et al., 2018). This process unfolds in four sequential steps. First, during identification, learners search for potential relationships between texts. Second, they create separate representations of the relevant information from each text independently. Third, through simultaneous relation, they draw explicit connections between the content of the different texts. Finally, during relational designation, learners classify the identified relationship, determining whether it reflects similarities, additions, or contradictions.

After identifying relationships between sources, learners engage in synthesis, defined as the integration of information from multiple texts into a coherent and comprehensive understanding. During synthesis, the information across multiple texts may align, complement, or contradict each other. When discrepancies arise, learners perform reconciliation, a sub-process of synthesis, which may take several forms: they may attempt to resolve conflicts between multiple texts, favor one text over another, or disregard the conflict entirely. Successful

synthesis marks the completion of the integration process (List & Alexander, 2019). Therefore, this model implies that learners may exhibit varying approaches to integrate multiple texts. Applied to the processing of multiple peer feedback messages, this may result in distinct patterns of peer feedback integration. These integration patterns are important for the design of CSCL environments, because without understanding them, current CSCL tools might inadvertently overload students or encourage superficial processing when assigning multiple peers.

The influence of objective feedback features and perceived feedback quality on integration

Recently, Hsiao and Rajagopal (2023) applied the Integrated Framework of Multiple Text Use to propose a design framework intended to support students in integrating multiple peer feedback messages within thesis writing circles. Although they did not test their model empirically, their work illustrates that theoretical insights from research on learning with multiple documents can provide a valuable foundation for understanding how students combine feedback from different peers. Building on this premise, the present study aims to empirically examine the processes by which students integrate multiple peer feedback messages and to investigate how the characteristics of received feedback influence this integration.

Based on prior research, we posit that the features of feedback messages students receive from peers can shape how they integrate them. Thereby, both objective feedback features and the perceived quality of the feedback are likely to play a role. Objective feedback features include the content and structural quality of the feedback (Nelson & Schunn, 2008; Wu & Schunn, 2020). Content quality refers to the accuracy and relevance of the feedback—that is, whether it correctly addresses the task. Structural quality, in contrast, concerns the presence of specific feedback elements, such as reminders of task requirements (feed up), evaluations of student performance (feed back), or suggestions for improvement (feed forward; Hattie & Timperley, 2007). Effective structural feedback is also expected to be clear, illustrative, and targeted at the aspects of the task that require attention, and it may include explanations or examples that clarify why certain aspects are correct or incorrect (Gielen & De Wever, 2015). Wu and Schunn (2020) found that explanations in feedback increase the likelihood of its implementation, and that this effect may be even stronger when learners receive multiple peer feedback messages. By contrast, Nelson and Schunn (2008) observed that summaries and concrete solution suggestions were more likely to prompt implementation than explanations. Despite these mixed findings, evidence suggests that objective feedback features play a critical role in guiding how students integrate feedback.

In addition to objective feedback features, the recipient's perception of the quality of the feedback they received may influence the integration process (Strijbos et al., 2021). Perceived feedback quality encompasses both the adequacy of feedback (e.g., fairness, usefulness, and acceptance; Strijbos et al., 2021), and subjective evaluations of specific feedback elements (e.g., feed up, feed back, feed forward; Hattie & Timperley, 2007). Strijbos et al. (2021) demonstrated that students' perceptions of fairness, usefulness, and acceptance predict their motivation to improve and their affective responses. Conversely, Huisman et al. (2018) found no effect of perceived feedback quality on peer feedback outcomes. Although findings are somewhat inconsistent, perceived feedback quality appears to influence how students engage with feedback.

Overall, previous research suggests that both objective feedback features and perceived feedback quality can affect subsequent processes on the part of the feedback recipient. However, most prior studies have focused on feedback implementation, whereas research on how learners integrate multiple peer feedback messages prior to implementation remains limited. We propose that both objective feedback features and perceived feedback quality are likely to influence feedback integration (Richter & Maier, 2017). For example, if one feedback message contains inaccurate information, while another is correct (content quality), the recipient needs to reconcile the conflict during integration, whereas integrating two accurate messages may require less effort. Similarly, if a student perceives one peer feedback message as useful and another as not, the latter may be ignored. Empirical studies testing these assumptions are currently lacking.

Research questions

The present study investigates how pre-service teachers integrate feedback they received from multiple peers and examines how objective feedback features (accuracy, relevance, implementability, structural feedback quality) and perceived feedback quality (adequacy, subjective quality) shape this integration process. Specifically, the study addresses the following questions:

(1) Can different integration patterns be distinguished empirically?

(2) Do objective feedback features and perceived peer feedback quality predict integration?

Methods

Sample and design

The study included $N = 100$ pre-service teachers from a German university, with a mean age of $M = 22.88$ years ($SD = 5.16$), who were on average in their fourth semester ($M = 4.47$, $SD = 1.01$), indicating that participants were approximately halfway through their teacher education program. Consistent with typical demographics of German pre-service teacher cohorts, 77% of the sample were female. The study was embedded within three regular courses, all of which covered identical content about educational psychology. Students engaged in a CSCL environment as part of these courses. This environment presented authentic case scenarios in the form of written descriptions of problematic classroom situations, corresponding analytical tasks, and relevant educational evidence. Participation in the peer feedback exercise was required for course credit; however, pre-service teachers could opt out of the use of their data for research purposes without any academic consequences, in which case their data were removed from the dataset. The sample analyzed in this study is a subset of a larger dataset that has been published elsewhere (Hornstein, Greisel & Kollar, 2025).

Digital peer feedback tool

The peer feedback process was conducted using the computer-supported peer feedback tool "getFeedback!", which enables instructors to provide targeted guidance for each phase of the peer feedback process, including the task performance, feedback provision, feedback reception, and revision stages. Within this tool, all feedback exchanges are anonymous, and the tool randomly assigns students to provide and receive feedback from different peers (Fey et al., 2023).

Task and procedure

The peer feedback process took place over four weeks, with participants completing all tasks remotely through the "getFeedback!" tool. In Week 1 (task performance phase), participants were presented with a written description of a hypothetical problematic teaching scenario and instructed to analyze it according to a five-step scheme for evidence-informed reasoning: problem identification, problem description, problem explanation, goal setting, and selection of actions (Greisel et al., 2022). To support their analysis, participants were provided with brief summaries of two theories from educational psychology: cognitive load theory (Sweller, 1999) and the ICAP model (Chi & Wylie, 2014). In Week 2 (feedback provision phase), each participant's analysis was randomly assigned to two peers, who were tasked with providing feedback on the assigned analyses. During Week 3 (feedback reception and revision phase), participants received their own analyses along with the two peer feedback messages and were instructed to integrate the feedback and revise their initial work accordingly. Finally, in Week 4, participants analyzed a second authentic teaching scenario, again using cognitive load theory and the ICAP model, but this time without peer feedback, to assess their domain-specific reasoning performance.

Instruments

Feedback Integration. To capture participants' integration processes, participants were asked to highlight the comments in the feedback they received from peers using different colors: green for similarities, yellow for additions, and red for contradictions between the two feedback messages. For this, students received a step-by-step instruction based on the Integrated Framework of Multiple Text Use (List & Alexander, 2019), demonstrated through a sample solution. First, participants read both feedback messages. Second, they marked similarities, additions, and contradictions using the color-coding scheme. Third, they revised their analyses based on these markings. Finally, in the fourth step, participants checked their markings again within their revised analyses.

Objective Feedback Features. Each feedback message was analyzed in terms of content quality (accuracy and relevance), structural quality, and implementability. Content and structural quality were coded using a theory-driven framework (Hattie & Timperley, 2007; Nelson & Schunn, 2008; Yu & Schunn, 2023). *Content quality* was assessed using the dimensions of accuracy and relevance (Nelson & Schunn, 2008). A feedback message was coded as accurate if its content was correct relative to the original case analysis, and relevant if it addressed all major deficiencies in the analysis. *Structural quality* was operationalized using eight criteria, such as comparisons between the initial solution and task requirements, feedback length, and level of detail. An exploratory factor analysis (based on the Kaiser-Guttman criterion, scree plot, and MAP test) indicated that all eight criteria loaded on a single factor. Factor scores derived from a confirmatory factor analysis were used to compute a composite structural quality score. Each feedback message was analyzed as a whole across all problem analyses. To validate the coding, two independent coders rated 10 % of the data, achieving sufficient to excellent interrater reliability (Gwet's $AC1 = 0.66–0.94$). Discrepancies were resolved through discussion, and the remaining 90% of messages were divided among the coders. Feedback was also coded for implementability;

 © ISLS

two coders independently rated 10% of the messages, achieving good agreement (Gwet's *AC1* = 0.83), after which the remaining data were equally divided and coded.

Perceived Feedback Quality. Perceived feedback quality was measured in terms of adequacy and subjective feedback quality. Adequacy (Peer 1: ω = .80; Peer 2: ω = .82) was assessed using Strijbos et al.'s (2021) Feedback Perception Scale on a nine-point Likert scale, covering three dimensions with three items each: fairness (e.g., "I would be satisfied with this feedback"), usefulness (e.g., "I would consider this feedback useful"), and acceptance (e.g., "I would accept this feedback"). Subjective feedback quality (Peer 1: ω = .71; Peer 2: ω = .82) was measured using 18 items derived from Hattie and Timperley's (2007) feedback dimensions, rated on a five-point Likert scale. This included Feed Up (e.g., "The feedback specifies the goal of the task"), Feed Back (e.g., "The feedback informs me whether I have met the requirements"), and Feed Forward (e.g., "The feedback indicates how I could expand my elaboration").

Statistical analyses

To address RQ1, we examined the number of student markings in each color category—similarities, additions, and contradictions—and conducted a hierarchical cluster analysis using the Ward method with Euclidean distances to identify distinct peer feedback integration patterns. To further characterize differences between the identified patterns, we performed separate ANOVAs for the frequencies of similarities, additions, and contradictions, followed by post-hoc Tukey HSD tests. For RQ2, we conducted multinomial logistic regression analyses, using the objective feedback features (accuracy, relevance, implementability, structural feedback quality) and the two perceived feedback quality measures (adequacy, subjective feedback quality) as predictors, and the integration patterns as outcome variable. All analyses were performed in R [4.3.2], with the alpha level set to 5%.

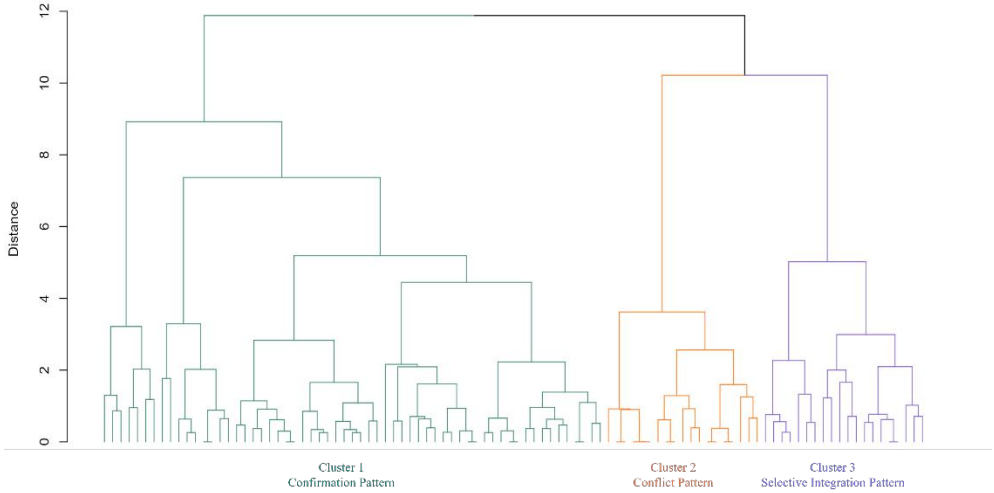
Results

RQ1: Can different integration patterns be distinguished empirically?

To explore distinct integration patterns (RQ1), we conducted a hierarchical cluster analysis using the number of markings learners made for similarities, additions, and contradictions in the feedback they received. All values were scaled to ensure comparability across categories. The analysis, based on the agglomerative coefficient (*AC* = 0.96), revealed three distinct clusters (see Figure 1).

Figure 1

Dendrogram of the Hierarchical Cluster Analysis Using Ward's Method



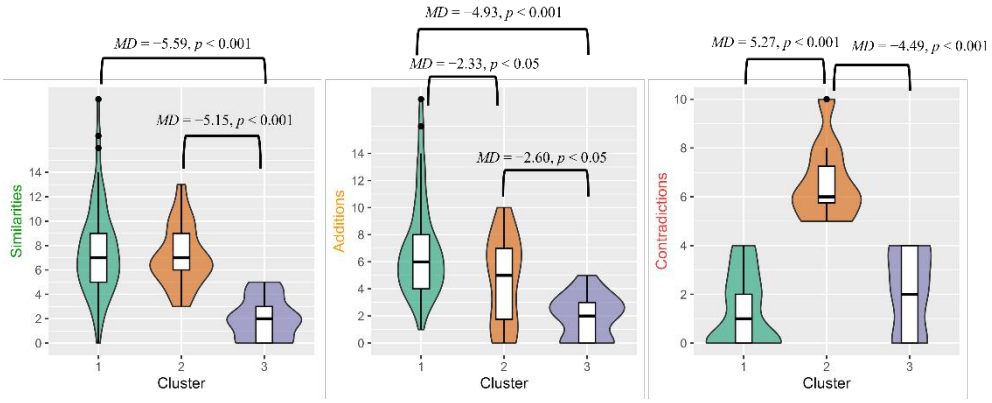
Cluster 1, which we labeled "confirmation pattern" and representing the majority of the sample (n = 61), included participants who highlighted a large number of similarities and additions. Cluster 2, which we termed the "conflict pattern", comprised participants who identified a high number of contradictions compared to the other clusters (n = 20). Cluster 3, which we referred to as "selective integration pattern", consisted of participants who marked feedback items rarely overall (n = 19). Violin plots in Figure 2 illustrate the frequencies of similarities, additions, and contradictions for each cluster, highlighting the differences visually. ANOVAs indicated significant differences among the clusters for similarities ($F(2,97) = 22.23$, $p < 0.001$, $\eta^2 = 0.31$),

 © ISLS

additions ($F(2,97) = 17.97$, $p < 0.001$, $\eta^2 = 0.27$), and contradictions ($F(2,97) = 92.37$, $p < 0.001$, $\eta^2 = 0.66$). Post-hoc Tukey HSD tests further clarified these differences: For similarities, the selective integration pattern identified fewer instances than both the confirmation pattern and the conflict pattern. Regarding additions, participants from the selective integration pattern marked less frequently than both participants from the confirmation and the conflict pattern. In addition, participants from the conflict pattern identified fewer additions than participants from the confirmation pattern. For contradictions, participants from the conflict pattern highlighted more than participants from the confirmation pattern, while participants from the selective integration pattern identified fewer contradictions than participants from the conflict pattern. In summary, these analyses demonstrate that the three clusters reflect robustly distinct integration patterns, differing in the extent to which learners recognize similarities, additions, and contradictions in peer feedback messages.

Figure 2

Violin Plots for Means of Frequencies of Similarities, Additions, and Contradictions per Cluster

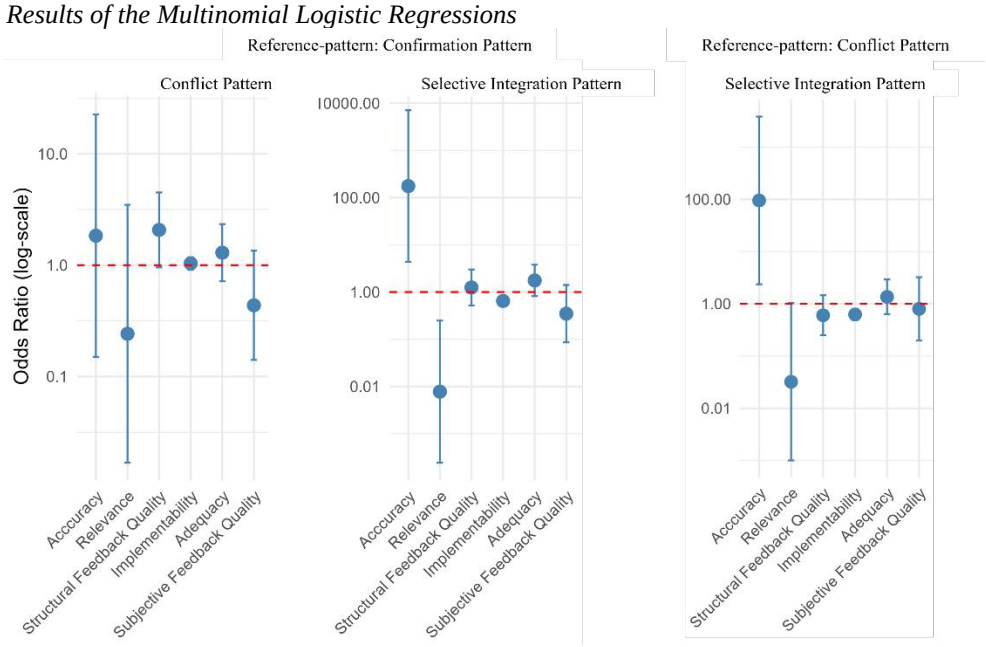


RQ2: Do objective feedback features and perceived peer feedback quality predict integration?

To examine how objective feedback features (accuracy, relevance, implementability, structural feedback quality) and perceived feedback quality (adequacy, subjective feedback quality) relate to participants' integration patterns (RQ2), we conducted a multinomial logistic regression, using the integration pattern as the criterion variable and the feedback features as predictors. The confirmation pattern and the conflict pattern were treated as reference categories, and model coefficients were interpreted as odds ratios (see Figure 3).

Figure 3

Results of the Multinomial Logistic Regressions



 © ISLS

Regarding objective feedback features, the results indicated that, compared to the confirmation pattern, membership in the selective integration pattern was significantly associated with accuracy ($OR = 175.57$, $p < .01$), relevance ($OR = 0.01$, $p < .01$), and implementability ($OR = 0.65$, $p < .001$). Specifically, a higher feedback accuracy increased the likelihood of participants exhibiting the selective integration pattern, whereas greater relevance and implementability decreased it. Additionally, when comparing the conflict pattern to the selective integration pattern, both accuracy ($OR = 95.49$, $p < .05$) and implementability ($OR = 0.63$, $p < .001$) were significant predictors. Again, higher accuracy and lower implementability were associated with a greater probability of belonging to the selective integration pattern.

In contrast, perceived feedback quality did not significantly differentiate between integration patterns, with all associated p-values exceeding .05.

Discussion

Computer-supported peer feedback is widely employed as a learning strategy in CSCL environments (e.g., Özbek et al., 2024; Zong et al., 2021). Nevertheless, questions remain regarding how and to what extent pre-service teachers integrate multiple peer feedback messages (RQ1), and the influence of objective feedback features and perceived feedback quality on this integration (RQ2). In the following, we provide a summary and interpretation of our findings in relation to these two research questions.

RQ1: Can different integration patterns be distinguished empirically?

Findings from the hierarchical cluster analysis revealed three distinct feedback integration patterns: the confirmation pattern, the conflict pattern, and the selective integration pattern. The confirmation pattern, which occurred most frequently, was characterized by numerous markings of similarities and additions between peer feedback messages. In contrast, the conflict pattern involved a higher frequency of contradictions identified by participants, while the selective integration pattern was marked by comparatively few overall annotations. These patterns closely mirror the synthesis and reconciliation processes outlined by List and Alexander (2019). Specifically, learners following the confirmation pattern seem to construct a synthesis by emphasizing convergent and complementary elements while overlooking inconsistencies. Those adopting the conflict pattern, however, appear to focus on resolving contradictions between feedback messages. Learners displaying the selective integration pattern seem to engage with only a limited number of feedback elements, selectively synthesizing specific aspects of the input.

Recognizing these distinct integration patterns offers promising opportunities for developing adaptive scaffolds within CSCL environments that use peer feedback. For instance, learners who demonstrate a confirmation pattern might benefit from prompts encouraging them to critically examine discrepancies between the peer feedback messages. Such guidance could promote deeper engagement with feedback, potentially enhancing both feedback implementation and learning outcomes (Richter & Maier, 2017). Conversely, learners following a conflict pattern could be supported with prompts emphasizing commonalities and complementarities between feedback and task solutions to mitigate possible frustration stemming from perceived criticism (Panadero et al., 2023).

Learners exhibiting a selective integration pattern may require special consideration. On the one hand, they might profit from more explicit scaffolding targeting all three integration processes, each of which can foster learning (Barzilai et al., 2018). On the other hand, it is also plausible that these learners already engage in mindful integration (Berndt et al., 2018), intentionally selecting only the feedback they deem most useful. In the latter case, additional prompting might not be necessary. Distinguishing between these possibilities would require further investigation, potentially employing qualitative approaches such as think-aloud protocols or eye-tracking (see van Meenen et al., 2023, for initial work in this area). Moreover, generative AI tools could prove valuable for detecting integration patterns and providing real-time adaptive support during feedback reception (Bauer et al., 2025). Future studies should explore the potential of these tools to enhance learners' engagement with peer feedback processes.

RQ2: Do objective feedback features and perceived peer feedback quality predict integration?

We also examined how the feedback participants received predicted the different integration patterns we identified. In doing so, we distinguished between objective feedback features—namely accuracy, relevance, implementability, and structural feedback quality—and perceived feedback quality, which included adequacy and subjective feedback quality. Our analyses revealed significant associations between objective feedback features and the observed integration patterns. Specifically, the accuracy, relevance, and implementability of the feedback

emerged as predictors of how participants engaged in integrating peer feedback. Participants who displayed the selective integration pattern tended to receive feedback that was highly accurate but less relevant and less implementable compared to those showing the other two patterns.

One possible explanation for this finding comes from a motivational perspective: when feedback lacks relevance or cannot be readily implemented, learners might be less inclined to highlight or annotate it, perceiving it as unlikely to improve their original work during revision (Strijbos et al., 2021). Interestingly, however, the selective integration pattern was also associated with receiving more accurate feedback. Compared to the confirmation pattern and conflict pattern, this could suggest that learners engaging in selective integration might be performing a form of mindful integration (Berndt et al., 2018). More accurate feedback may demand less cognitive effort to process, leading learners to agree with it and only mark these specific segments. In such cases, when the feedback aligns closely with learners' own understanding, they might accept it implicitly and proceed with selective annotations.

Surprisingly, we found no significant relationship between perceived feedback quality and integration patterns. This was unexpected to some extent, as prior studies suggest that learners' agreement with and perceptions of feedback quality play a substantial role in whether they implement feedback (Strijbos et al., 2021). However, it is also consistent with existing research showing that the perceived feedback quality does not influence its implementation (Huisman et al., 2018). A potential explanation could be that when objective feedback quality is consistently high, its influence may overshadow learners' subjective perceptions, rendering perceived quality less relevant to how feedback is integrated (Nelson & Schunn, 2008; Gielen & De Wever, 2015).

Overall, these findings emphasize the importance of maintaining high objective feedback quality in peer feedback contexts. When the accuracy, relevance, and implementability of feedback are ensured, learners may be more likely to engage productively in the integration process. Future research should further explore these mechanisms to clarify how different dimensions of feedback quality jointly shape feedback processing and integration behavior.

Limitations

Of course, this study has several limitations that should be acknowledged. First, the integration patterns identified in our study may be closely tied to the specific learning content and context used. To make more generalizable claims about how learners integrate multiple peer feedback messages, additional research is needed across different subject domains and learning environments to test the robustness and transferability of these patterns. Second, although participants were explicitly instructed to visualize their integration processes by using color-coded markings, it remains uncertain how diligently and consistently participants followed these instructions. To gain deeper insights, future research should aim to capture the underlying cognitive mechanisms that guide learners in identifying similarities, additions, and contradictions when processing feedback from multiple peers. Third, there are likely further influencing factors that were not addressed in this study but may play a crucial role in how learners integrate peer feedback. These could include variables such as learners' comprehension of the feedback, their motivation, or individual characteristics that shape how they process and apply feedback information. Future work should therefore seek to identify and model such individual-level predictors to provide a more comprehensive understanding of feedback integration processes.

Conclusions

Despite these limitations, the findings of this study provide empirical evidence that distinct patterns of peer feedback integration can be systematically identified in CSCL settings. Moreover, the results show that the likelihood of specific integration patterns is influenced by certain objective feedback features. This understanding offers valuable implications for the design of targeted, pattern-specific interventions to support learners in effectively integrating multiple peer feedback messages. Such interventions could be enhanced through generative AI tools or learning analytics systems capable of providing real-time, adaptive feedback support (Bauer et al., 2025). Exploring these possibilities in future research could significantly advance the practical application of peer feedback processes in digital learning environments.

 © ISLS

References

Alqassab, M., Strijbos, J.-W., & Ufer, S. (2018). Training peer-feedback skills on geometric construction tasks: Role of domain knowledge and peer-feedback levels. *European Journal of Psychology of Education, 33*(1), 11–30. https://doi.org/10.1007/s10212-017-0342-0

Barzilai, S., Zohar, A. R., & Mor-Hagani, S. (2018). Promoting integration of multiple texts: A review of instructional approaches and practices. *Educational Psychology Review, 30*(3), 973–999. https://doi.org/10.1007/s10648-018-9436-8

Bauer, E., Sailer, M., Niklas, F., Greiff, S., Sarbu-Rothsching, S., Zottmann, J. M., Kiesewetter, J., Stadler, M., Fischer, M. R., Seidel, T., Urhahne, D., Sailer, M., & Fischer, F. (2025). AI-based adaptive feedback in simulations for teacher education: An experimental replication in the field. *Journal of Computer Assisted Learning, 41*(1), Article e13123. https://doi.org/10.1111/jcal.13123

Berndt, M., Strijbos, J.-W., & Fischer, F. (2018). Effects of written peer-feedback content and sender's competence on perceptions, performance, and mindful cognitive processing. *European Journal of Psychology of Education, 33*(1), 31–49. https://doi.org/10.1007/s10212-017-0343-z

Chen, T. (2016). Technology-supported peer feedback in ESL/EFL writing classes: A research synthesis. *Computer Assisted Language Learning, 29*(2), 365–397. https://doi.org/10.1080/09588221.2014.960942

Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist, 49*(4), 2019–2243. https://doi.org/10.1080/00461520.2014.965823

Cho, K., & Schunn, C. D. (2007). Scaffolded writing and rewriting in the discipline: A web-based reciprocal peer review system. *Computers & Education, 48*(3), 409–426. https://doi.org/10.1016/j.compedu.2005.02.004

Dmoshinskaia, N., Gijlers, H., & Jong, T. de (2021). Giving feedback on peers' concept maps in an inquiry learning context: The effect of providing assessment criteria. *Journal of Science Education and Technology, 30*(3), 420–430. https://doi.org/10.1007/s10956-020-09884-y

Farahani, A. A. K., Nemati, M., & Nazari Montazer, M. (2019). Assessing peer review pattern and the effect of face-to-face and mobile-mediated modes on students' academic writing development. *Language Testing in Asia, 9*(1), 1–24. https://doi.org/10.1186/s40468-019-0094-7

Fey, C.-C., Wekerle, C., Beckmann, I., Schröder, A., & Kollar, I. (2023). Agile Methoden in Entwicklungsprojekten zur Innovation digitaler Hochschullehre [Agile Methods in Development Projects for Innovation in Digital Higher Education]. *Zeitschrift für Hochschulentwicklung [Journal of Higher Education Development], 18*(3), 275–294. https://doi.org/10.21240/zfhe/18-03/14

Gielen, M., & De Wever, B. (2015). Structuring peer assessment: Comparing the impact of the degree of structure on peer feedback content. *Computers in Human Behavior*, 52, 315–325. https://doi.org/10.1016/j.chb.2015.06.019

Greisel, M., Wekerle, C., Wilkes, T., Stark, R., & Kollar, I. (2022). Pre-service teachers' evidence-informed reasoning: Do attitudes, subjective norms, and self-efficacy facilitate the use of scientific theories to analyze teaching problems? *Psychology Learning & Teaching*, *22*(1), 20–38. https://doi.org/10.1177/14757257221113942

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112. https://doi.org/10.3102/003465430298487

Hornstein, J., Greisel, M., & Kollar, I. (2025). Promoting pre-service teachers' evidence-informed reasoning through peer-feedback: Effects of a feedback provision scaffold and a feedback integration scaffold. *Instructional Science*, 1–27. https://doi.org/10.1007/s11251-025-09745-4

Hornstein, J., Keller, M. V., Greisel, M., Dresel, M., & Kollar, I. (2025). Enhancing the peer-feedback process through instructional support: A meta-analysis. *Educational Psychology Review*, *37*(2). https://doi.org/10.1007/s10648-025-10017-3

Hsiao, Y. P., & Rajagopal, K. (2023). Support student integration of multiple peer feedback on research writing in thesis circles. In O. Noroozi & B. de Wever (Eds.), *Social Interaction in Learning and Development. The power of peer learning: Fostering students' learning processes and outcomes* (1st ed. 2023, pp. 47–70). Springer International Publishing; Imprint Springer. https://doi.org/10.1007/978-3-031-29411-2_3

Huisman, B., Saab, N., van Driel, J., & van den Broek, P. (2018). Peer feedback on academic writing: undergraduate students' peer feedback role, peer feedback perceptions and essay performance. *Assessment & Evaluation in Higher Education, 43*(6), 955–968. https://doi.org/10.1080/02602938.2018.1424318

Kollar, I., & Fischer, F. (2010). Peer assessment as collaborative learning: A cognitive perspective. Learning and Instruction, 20(4), 344–348. https://doi.org/10.1016/j.learninstruc.2009.08.005

Li, L., Liu, X., & Steckelberg, A. L. (2010). Assessor or assessee: How student learning improves by giving and receiving peer feedback. *British Journal of Educational Technology, 41*(3), 525–536. https://doi.org/10.1111/j.1467-8535.2009.00968.x

List, A., & Alexander, P. A. (2019). Toward an Integrated Framework of Multiple Text Use. *Educational Psychologist, 54*(1), 20–39. https://doi.org/10.1080/00461520.2018.1505514

Lundstrom, K., & Baker, W. (2009). To give is better than to receive: The benefits of peer review to the reviewer's own writing. *Journal of Second Language Writing, 18*(1), 30–43. https://doi.org/10.1016/j.jslw.2008.06.002

Nelson, M. M., & Schunn, C. D. (2008). The nature of feedback: How different types of peer feedback affect writing performance. *Instructional Science, 37*(4), 375–401. https://doi.org/10.1007/s11251-008-9053-x

Nicol, D., & McCallum, S. (2022). Making internal feedback explicit: exploiting the multiple comparisons that occur during peer review. *Assessment & Evaluation in Higher Education, 47*(3), 424–443. https://doi.org/10.1080/02602938.2021.1924620

Nicol, D., Thomson, A., & Breslin, C. (2014). Rethinking feedback practices in higher education: A peer review perspective. *Assessment & Evaluation in Higher Education, 39*(1), 102–122. https://doi.org/10.1080/02602938.2013.795518

Özbek, T., Daumiller, M., Roshany-Tabrizi, A., Mömke, T., & Kollar, I. (2024). Friendship or feedback? – Relations between computer science students' goals, technology acceptance, use of an online peer feedback tool, and learning. *Computers in Human Behavior Reports, 16*, 100540. https://doi.org/10.1016/j.chbr.2024.100540

Panadero, E., & Alqassab, M. (2019). An empirical review of anonymity effects in peer assessment, peer feedback, peer review, peer evaluation and peer grading. *Assessment & Evaluation in Higher Education, 44*(8), 1253–1278. https://doi.org/10.1080/02602938.2019.1600186

Panadero, E., Alqassab, M., Fernández Ruiz, J., & Ocampo, J. C. (2023). A systematic review on peer assessment: intrapersonal and interpersonal factors. *Assessment & Evaluation in Higher Education*, 1053–1075. https://doi.org/10.1080/02602938.2023.2164884

Richter, T., & Maier, J. (2017). Comprehension of multiple documents with conflicting information: A two-step model of validation. *Educational Psychologist, 52*(3), 148–166. https://doi.org/10.1080/00461520.2017.1322968

Strijbos, J.-W., Pat-El, R., & Narciss, S. (2021). Structural validity and invariance of the Feedback Perceptions Questionnaire. *Studies in Educational Evaluation*, 68, Article 100980. https://doi.org/10.1016/j.stueduc.2021.100980

Sweller, J. (1999). *Instructional Design in Technical Areas*. ACER Press.

van Meenen, F., Masson, N., Catrysse, L., & Coertjens, L. (2023). Taking a closer look at how higher education students process and use (discrepant) peer feedback. *Learning and Instruction*, 84, 101711. https://doi.org/10.1016/j.learninstruc.2022.101711

van Popta, E., Kral, M., Camp, G., Martens, R. L., & Simons, P. R.-J. (2017). Exploring the value of peer feedback in online learning for the provider. *Educational Research Review*, 20, 24–34. https://doi.org/10.1016/j.edurev.2016.10.003

Vygotsky, L. S., Cole, M., Jolm-Steiner, V., Scribner, S., & Souberman, E. (1978). *Mind in society: The development of higher psychological processes.* Harvard University Press. https://doi.org/10.2307/j.ctvjf9vz4

Wu, Y., & Schunn, C. D. (2020). When peers agree, do students listen? The central role of feedback quality and feedback frequency in determining uptake of feedback. *Contemporary Educational Psychology*, 62, 101897. https://doi.org/10.1016/j.cedpsych.2020.101897

Yu, Q., & Schunn, C. D. (2023). Understanding the what and when of peer feedback benefits for performance and transfer. *Computers in Human Behavior*, 147, 107857. https://doi.org/10.1016/j.chb.2023.107857

Zong, Z., Schunn, C. D., & Wang, Y. (2021). What aspects of online peer feedback robustly predict growth in students' task performance? *Computers in Human Behavior*, 124, Article 106924, 1–13. https://doi.org/10.1016/j.chb.2021.106924

Acknowledgments

This work was supported by the Stiftung Innovation in der Hochschullehre (project "Facilitating Competence Development through Authentic, Digital, and Feedback-Based Teaching-Learning Scenarios", grant number FBM2020).

International Society of the Learning Sciences

Student, Hybrid, or AI?: Examining Authorial Roles and Network Structures in GenAI-Supported Collaborative Learning

Shaoming Chai, Aberdeen institute of Data Science and Artificial Intelligence, South China Normal University, charmingchai@m.scnu.edu.cn
Zhenhai He, Guangdong University of Science and Technology, Guangdong Nanbo Institute of Education, zhenhai@m.scnu.edu.cn
Tongwu Jiang, School of Information Technology in Education, South China Normal University, jiangtongwu@m.scnu.edu.cn

Abstract: As generative AI (GenAI) reshapes collaborative learning, understanding how authorial roles interact within evolving network structures is crucial. This exploratory case study explored authorial roles and social network structures in a Knowledge Forum-based computational thinking course. We analyzed 527 discussion notes from 54 undergraduates using Social Network Analysis (SNA) and deep learning classification. Notes were categorized into four roles: pure student (Stu), direct GenAI outputs, Student-GenAI (SG) hybrids, and Student-GenAI-Student (SGS) iterations. Results showed student-only contributions dominated (73.8%), with GenAI-involved notes at 26.2%. Network analysis revealed Stu roles as primary information hubs with high centrality, while hybrid SG/SGS roles were less central than expected. Findings suggest that without explicit guidance, students remain core connectors and AI-facilitated knowledge brokerage is underutilized. Educational practice must intentionally scaffold hybrid roles to foster balanced, integrated knowledge flow in GenAI-supported collaborative learning.

Introduction

In the evolving educational landscape, collaborative learning has emerged as a cornerstone for developing essential 21st-century skills, including computational thinking (CT) and problem-solving. Frameworks such as Knowledge Building (KB) emphasize the collective advancement of ideas within a community, where students engage in authentic discourse to refine and expand shared knowledge (Scardamalia & Bereiter, 2006). Platforms such as the Knowledge Forum (KF) facilitate this process by enabling students to post notes, discuss solutions, and iteratively build upon each other's contributions, creating a dynamic environment for idea exchange (Bereiter & Scardamalia, 2014). However, traditional collaborative settings often face limitations in scalability and personalization, prompting the exploration of technological integration to enhance engagement and outcomes.

The emergence of Generative AI (GenAI) tools like ChatGPT offers new opportunities for collaborative learning. GenAI acts as a virtual collaborator, generating ideas and providing feedback, thereby extending human capabilities in knowledge co-construction (Chan & Hu, 2023). Studies have shown that GenAI enhances creativity and problem-solving in group tasks by reducing cognitive load and fostering diverse perspectives (Wei et al., 2025). However, its integration raises questions about authorship and the balance between human and AI contributions, particularly in cases where students may not be able to distinguish their ideas from AI-generated content.

This study examined GenAI-supported collaborative learning within a KF-based course aimed at cultivating CT. As GenAI assumes increasingly diverse roles in educational contexts—evolving from a passive information provider to an active co-constructor of knowledge—understanding the dynamics of human-AI collaboration becomes crucial (Chan & Hu, 2023; Lee et al., 2023). In this learning environment, students collaborated on problem-solving tasks while incorporating GenAI interactions into their discussion notes. To capture the varying degrees of human-AI integration, pedagogical scaffolds were designed to differentiate four distinct authorial roles: pure student (Stu) contributions representing independent human thought, direct GenAI outputs showcasing AI-generated content, Student-GenAI (SG) hybrids reflecting initial human-AI integration, and refined Student-GenAI-Student (SGS) iterations demonstrating critical engagement with AI outputs. Owing to the varying scaffold proficiency, deep learning methods were employed to classify discourse into these four roles, followed by Social Network Analysis (SNA) to uncover interaction patterns. To guide this investigation, the following research questions were posed:

RQ1: What is the distribution of the four distinct authorial roles—Stu, GenAI, SG, and SGS—within the collaborative discourse in this AI-supported collaborative learning environment?

 © ISLS

RQ2: What typical interaction network structures emerge within student groups when considering the different authorial roles adopted by students?

RQ3: What structural roles (e.g., centrality, brokerage) do author-role nodes with different authorial role configurations play within the collaborative network? Specifically, do nodes representing Student-GenAI-Student author role combinations occupy critical bridging positions in the KB network?

Literature review

Collaborative learning and KB

Collaborative learning theories highlight the importance of social interaction in acquiring knowledge, wherein learners co-construct understanding through dialogue and shared activities. KB, a framework by Scardamalia and Bereiter (2006), views education as collective idea improvement, not mere transmission, emphasizing idea diversity, improvable ideas, and synthesis. In platforms like KF, this appears as threaded discussions where students post notes, reference peers' ideas, and refine solutions, fostering computational thinking in problem-solving (Bereiter & Scardamalia, 2014). Studies have shown that such environments enhance metacognitive skills and group efficacy, although challenges include ensuring equitable participation and distinguishing individual contributions (Grunspan et al., 2014).

The integration of AI into these settings extends traditional collaborative paradigms through hybrid human-AI discourse. In KB scenarios, AI serves as a scaffold for idea generation, aligning with the principles of authoritative sources (Lee et al., 2023). This hybridization complicates authorial roles, as students may paste AI outputs into their notes, blurring the lines between original and generated content. Deep learning classification techniques parse such discourses by categorizing contributions based on linguistic patterns (Wang et al., 2023). This reveals that SG hybrids predominated in creative tasks, suggesting that AI's role is to amplify ideas while risking over-reliance.

Genai in educational environment

GenAI has transformed educational practices by enabling personalized support in learning environments. Tools such as ChatGPT facilitate content creation, brainstorming, and feedback, thereby enhancing student engagement (Chan & Hu, 2023). In higher education, GenAI supports discourse by summarizing discussions, generating recommendations, and synthesizing knowledge, aligning with knowledge creation goals (Chen et al., 2023; Cress & Kimmerle, 2023). Experimental studies have shown that GenAI boosts collaborative problem-solving and creativity, as seen in digital storytelling, where tools like Midjourney and Runway aid visual and narrative development (Wei et al., 2025). However, perceptions vary; students appreciate efficiency but express concerns about accuracy, ethics, and diminished critical thinking (Chan & Hu, 2023).

In KB contexts, GenAI acts as a co-constructor, prompting reflection and refinement, although its lack of understanding requires human oversight (Lee et al., 2023). Frameworks such as Know-Build-Critique guide ethical integration, emphasizing AI literacy (Chen et al., 2023). Authorial roles emerge as critical, with classifications of pure students versus AI-augmented input revealing patterns in which refined SGS contributions foster discourse (Cress & Kimmerle, 2023). Challenges include cognitive offloading, where reliance on AI weakens personal development, requiring scaffolding for active refinement.

SNA in learning environments

SNA provides a robust method for mapping interactions in educational networks, revealing structures such as centrality and brokerage that indicate influence and bridging roles (Grunspan et al., 2014). In AI-supported learning, SNA uncovers human-AI dynamics by clustering participants based on interaction patterns and epistemic networks to assess community sense and performance (Wang et al., 2023). In GenAI contexts, SNA identifies emergent structures where hybrid roles, such as SGS, occupy bridging positions, facilitating knowledge flow across groups (Feng, 2025). Network methods analyze small-group interactions, providing evidence of how GenAI influences discourse patterns and equity (Feng, 2025). Gaps persist in applying SNA to classify authorial roles in KF, where deep learning aids in delineating nodes, potentially revealing whether refined hybrids enhance centrality and brokerage in computational thinking tasks. Despite these advancements, the existing literature lacks integrated analyses of authorial role distributions and network structures in GenAI-supported KB. This study aims to address this gap by classifying discourse via deep learning, applying SNA to explore patterns, and elucidating the structural significance of hybrid roles in collaborative networks.

Methodology

This study employed an exploratory case study methodology (Yin, 1981) with quantitative social network analysis (SNA) and deep learning techniques for automated text classification to investigate AI-supported collaborative learning environments. The integration of generative AI into collaborative learning represents a phenomenon in which the boundaries between human and artificial agents are increasingly blurred. The case study design allowed for a contextual examination of human-AI interaction dynamics, while SNA provided an analytical framework to visualize collaborative networks. Deep learning-based text classification enables the systematic analysis of educational discourse by connecting qualitative understanding with quantitative modeling.

Participants

The study involved 54 second- and third-year undergraduate students from a key public university in China, enrolled in the general education course "Computational Thinking and Problem-Solving." The participants represented diverse academic backgrounds, including software engineering, artificial intelligence, cultural industry management, and financial management, ensuring disciplinary heterogeneity within the cohort. Students were organized into 10 interdisciplinary groups of 5-6 members each to facilitate collaborative problem-solving activities.

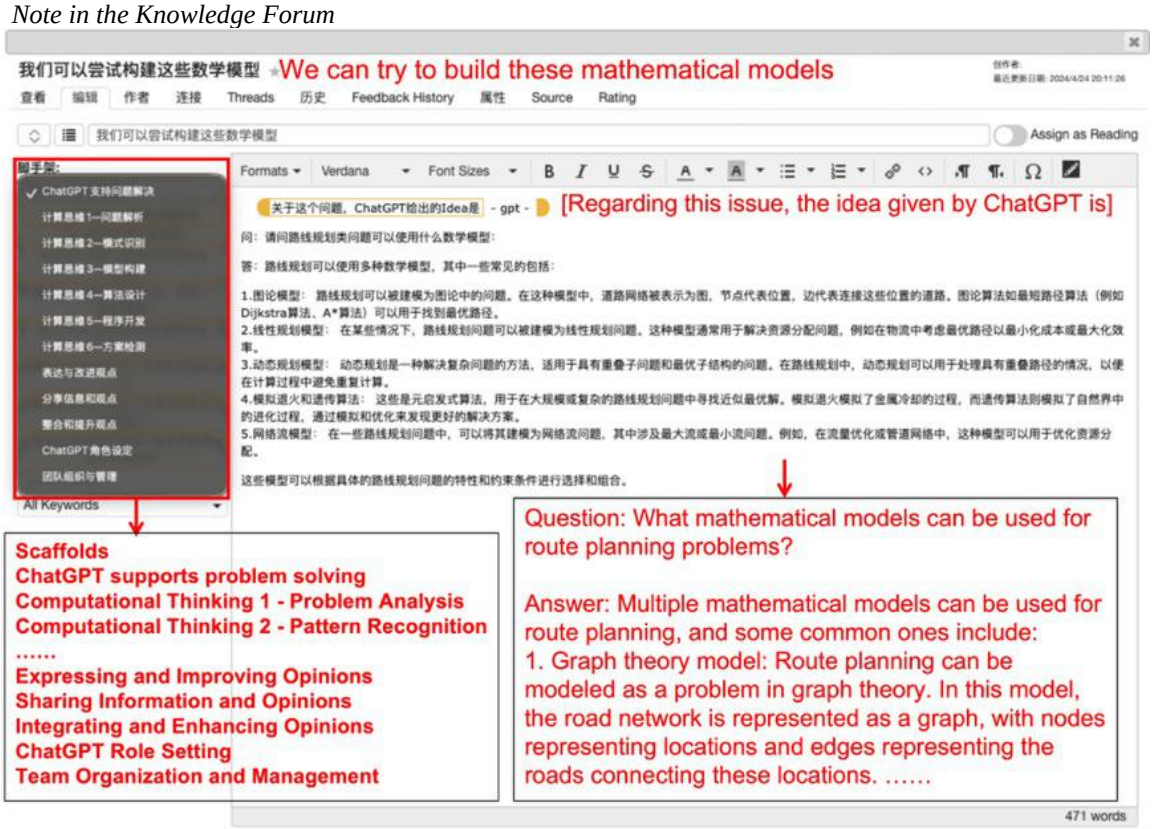
The learning environment

The study was conducted during the Spring 2025 semester over a 12-week period within a Problem-Based Learning (PBL) framework guided by knowledge-building principles. All collaborative activities occurred on the KF platform, augmented by two core pedagogical components:

Pedagogical Support for GenAI Use (see Figure1): Students were encouraged to integrate various external GenAI tools (ChatGPT, DeepSeek, Doubao) into their inquiry processes. Specialized scaffolds within the KF platform structured these interactions, including prompts such as "The explanation for this concept from ChatGPT is...", "I think ChatGPT's answer/explanation is still lacking...", and "Here is how I evaluate the accuracy and reliability of ChatGPT's proposal."

Figure 1

Note in the Knowledge Forum



Explicit Human-AI Interaction Training: Dedicated instruction focused on effective GenAI interaction strategies, including prompt engineering, multi-turn dialogue engagement, and critical evaluation of AI-generated responses.

Data collection and analysis

The primary data source consisted of 527 valid discussion notes generated on the KF platform during the group activities. Each note included metadata such as the group ID, note ID, note body, scaffold(s) used, and author type. Notes unrelated to core collaboration (e.g., reflections, final reports, videos or PPTs) were excluded. Data preprocessing involved cleaning the text content, handling Chinese word segmentation using libraries such as Jieba, and preparing features for classification, including the note text and associated scaffolds. Data analysis combined qualitative content analysis (QCA), deep learning–based text classification, and SNA to address the complex interactional dynamics of GenAI-supported collaborative learning. The analysis was conducted in two stages. First, Authorial Role Classification (ARC) was conducted to identify the degree of human–AI integration in the notes. Second, group interaction networks were constructed and analyzed to reveal their structural patterns and role functions.

Authorial role classification: A deep learning approach

To address RQ1, we developed a four-category ARC scheme from the study data, grounded in CSCL discourse analysis principles (Weinberger & Fischer, 2006), to identify varying levels of human–AI integration in the notes (Table 1): Stu (purely student-authored content), GenAI (verbatim AI output), SG (records of human–AI interaction), and SGS (student critique or integration following AI content).

Table 1

Four-Category ARC Scheme

Label ID	Category	Definition	Typical Characteristics
1	Student (Stu)	The note's content consists entirely of the student's original ideas, viewpoints, analysis, or reflections, without any content directly copied from GenAI.	Personalized expression, independent thinking, critical reflection.
2	GenAI	The student has directly and completely copied an answer generated by GenAI, with no personal additions or commentary.	Robotic tone, fluent language but lacking a personal student perspective.
3	Student-GenAI (SG)	The student has copied a log of their question-and-answer session with GenAI, either in full or in part, without further elaboration or integration.	Presented as a Q&A transcript; shows the process of interaction but lacks synthesis or innovation based on that process.
4	Student-GenAI-Student (SGS)	The student has pasted content from GenAI and then added their own in-depth thinking, critique, integration, or novel ideas.	The note typically begins with a quote or summary from AI, followed immediately by the student's own analysis, critique, or proposal.

The classification procedure began with manual annotation. From the 527 valid notes, we drew a stratified random sample of 220. Two researchers independently coded the data, and interrater reliability assessed via Cohen's κ indicated excellent agreement (κ = 0.811). Agreement was highest for the Student category, whereas Student–GenAI and Student–GenAI–Student showed minor confusion, primarily because of fuzzy category boundaries. The annotations further showed that Student accounted for 56.4% of the total, while GenAI was the least frequent at 11.4%. Next, we used the annotated data to train a bidirectional encoder representation from transformers (BERT) classifier. The input concatenated the scaffold type and note body as follows:

$$Input_{seq} = [CLS] \oplus (Scaffold_{type}) \oplus [SEP] \oplus Note_{body} \oplus [SEP]$$

where is $[CLS]$the classification token that aggregates the sequence representation, $[SEP]$separates segments, $Scaffold_{type}$ the tokenized scaffolding type, and $Note_{body}$ is the tokenized note content. The $\oplus$ operator denotes concatenation, which sequentially joins these components into a unified input sequence. This design allows the model to capture cross-segment semantic relations (referred to as "multimodal semantic relations" in this study). On the test set, the model achieved an accuracy of 0.818 and a macro-averaged F1 of 0.668. The performance was best for the Student class, whereas the recall for Student–GenAI–Student was lower. Finally, two researchers manually reviewed 49 instances with a prediction confidence below 0.9; the final κ reached 0.899, ensuring classification quality across the full dataset.

Group-level network structure analysis

To address RQ2, we computed group-level structural metrics, including network density (reflecting the tightness of interactions) and degree centralization (capturing centralized exchange patterns), and used a visualization of the interaction distribution to identify decentralized, peer-driven, or AI-centered structures (Wasserman & Faust, 1994). Metrics were calculated and layouts rendered with Python's NetworkX library, revealing emergent network patterns under different authorial role dominance.

Node-level centrality and brokerage role analysis

With respect to RQ3, we computed individual-level centrality measures in the directed interaction network: degree centrality (with in-degree and out-degree assessing influence and proactiveness, respectively), betweenness centrality (identifying brokerage roles), and closeness centrality (indicating the efficiency of information reach) (Chen & Poquet, 2022). We then compared the mean centrality scores across student-led authorial role types, with particular attention to whether the SGS role occupied key knowledge-broker positions. Statistical significance was tested using the Kruskal–Wallis test.

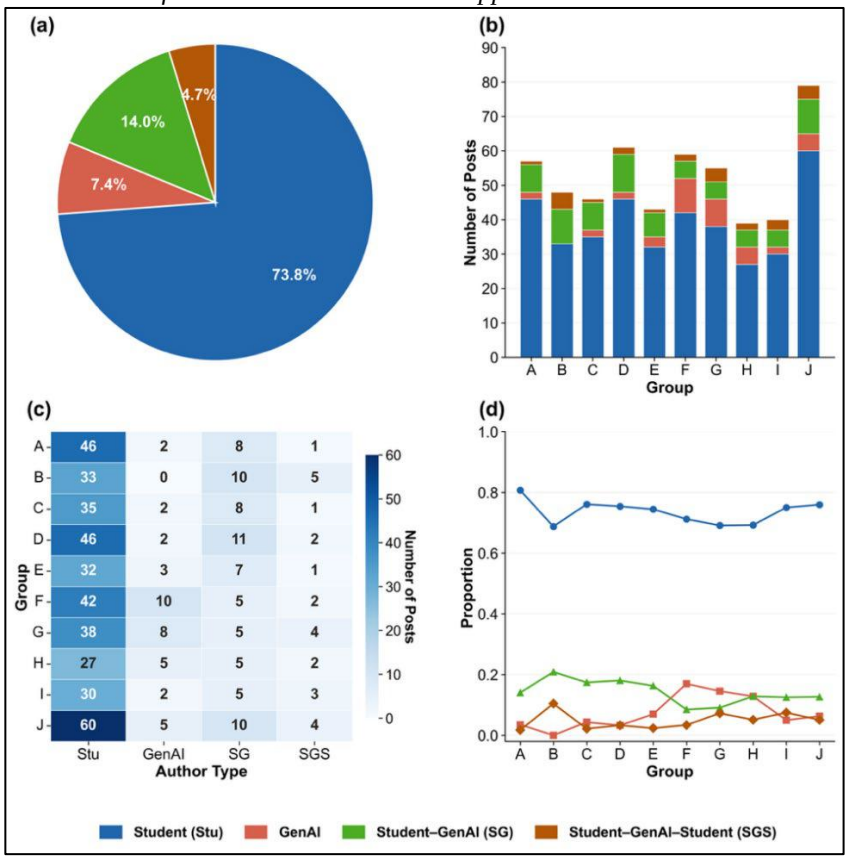
Result and discussion

RQ1: Distributional characteristics of roles in collaborative discourse

The distribution of the four authorial roles was markedly non-uniform (Figure 2a). Student-authored notes dominated, accounting for 73.8% of all contributions, whereas GenAI-involved notes comprised 26.2%. Students thus engaged primarily in independent sense-making, with AI contributions present but far from dominant. This pattern is consistent with recent findings. Wei et al. (2025), for example, reported that groups using GenAI tools in a digital storytelling experiment achieved significant gains in collaborative problem-solving and team creativity, yet human contributions remained central; GenAI enhanced efficiency and ideation rather than replacing student input. Together, these results suggest that GenAI functions as a cognitive augmentation tool in collaborative settings, extending rather than supplanting human reasoning and thereby reducing the risk of cognitive offloading.

Figure 2

Distribution of Authorial Roles in Genai-Supported Collaborative Notes



(a) overall role distribution, (b) role distribution by group, (c) heatmap of role counts by group, (d) trends in role proportions by group

 © ISLS

Between-group differences further highlighted this heterogeneity (Figures 2b–2c). For instance, the GenAI role was entirely absent in some groups (e.g., Group B) but reached 16.9% in others (e.g., Group F). SGS had the lowest prevalence across all groups (mean, 4.7%). This may be related to the scaffolds' design, suggesting that students face cognitive challenges when attempting to integrate AI outputs more deeply. Survey evidence by Ruiz-Rojas et al. (2024) indicates that although 87% of university students are familiar with GenAI tools, only 64% believe these tools substantially enhance critical thinking—consistent with the low share of SGS and potentially attributable to unresolved issues of bias and ethics that complicate effective integration in collaboration. Similarly, Zhai (2024) argues for a shift in teacher roles from observers to innovators, underscoring the need for ongoing professional development in the GenAI era to help students overcome these challenges and promote equitable AI use across groups.

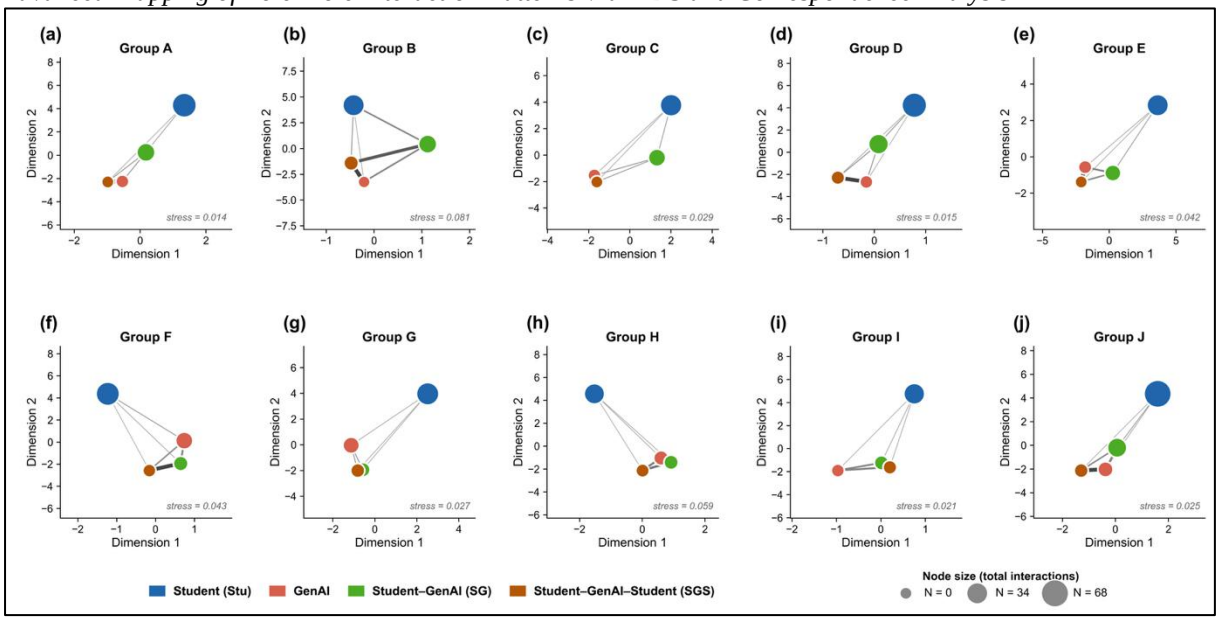
From the proportional trend plot (Figure 2d), the share of Stu remained relatively stable across groups (≈ 0.70 – 0.80), whereas SG varied the most, peaking at ~0.21 in Group B and falling to ~0.08 in Group F. This trend reveals the dynamic nature of GenAI involvement, which is potentially shaped by within-group interaction patterns or task design. Prior work supports this view; for instance, in GenAI-supported knowledge construction, student role allocations may be influenced by chatbots, yet the effects are often statistically non-significant and more salient at the perceptual level (Gong et al., 2024). Overall, the distribution indicates an assistive—not substitutive—positioning of GenAI in collaborative discourse, providing a diverse role landscape for subsequent network analyses.

RQ2: Identifying prototypical interaction-network structures

SNA results revealed diverse interaction networks in GenAI-supported collaborative learning. The mean network density was 0.408 and degree centralization was 0.367; most groups achieved full connectivity (connectivity = 1), as shown in Figures 3a, 3b, and 3e. This moderate density indicates interactions that avoid isolation while preventing information overload (Feng, 2025). Groups A and J showed densities of 0.583, forming grid-like structures that reflect frequent exchanges among roles—consistent with Wei et al. (2025), who found GenAI tools enhanced CPS and team creativity through increased interaction without displacing human contributions. Conversely, Group B showed lower density (0.333) and network fragmentation, suggesting interactional atomization due to over-reliance on GenAI or insufficient scaffolding (Zhang & Wang, 2025). This fragmentation risk is higher in AI-centered patterns, where students interact more with chatbots than with peers, dampening the peer-to-peer knowledge flow (Feng, 2025).

Figure 3

Advanced Mapping of Role–Role Interaction Patterns Via MDS and Correspondence Analysis

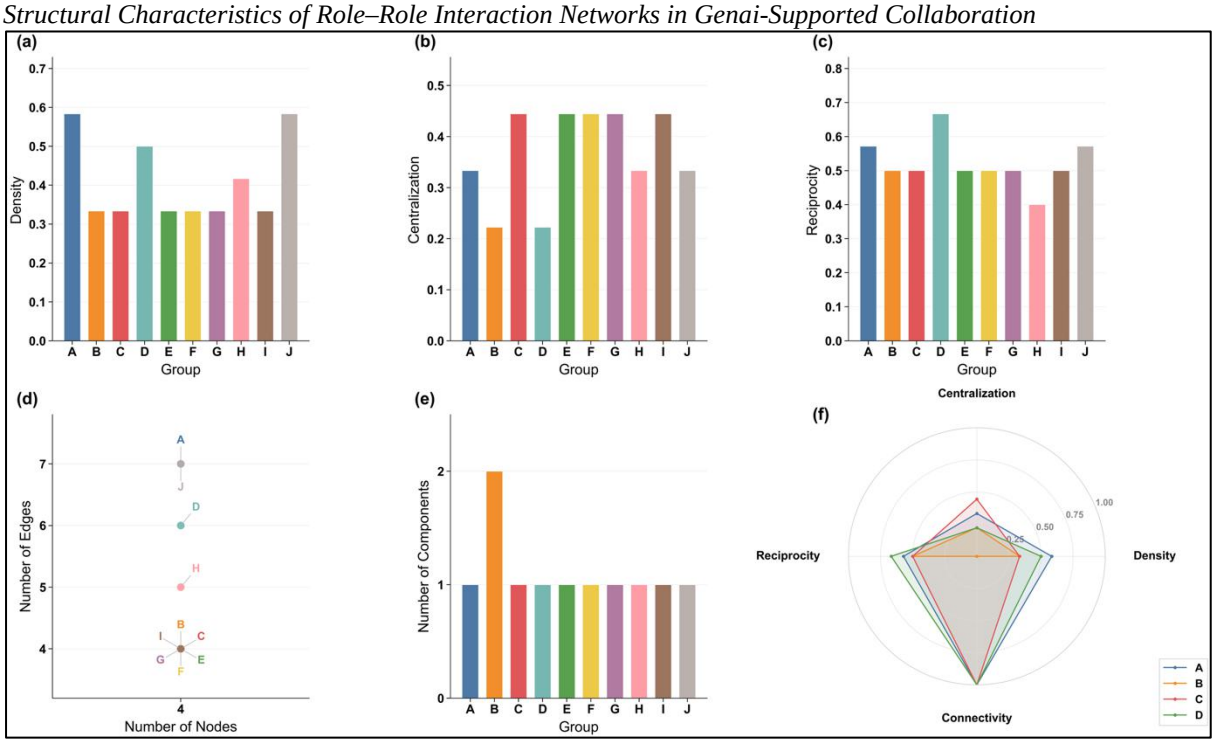


Reciprocity analysis (Figure 4c) further quantified the symmetric exchanges. High-reciprocity groups such as Group D (0.667) displayed balanced knowledge exchange, aligning with Zhang & Wang (2025), who found GenAI effective for information processing and critical reflection but less supportive of social interaction—

potentially leaving low-reciprocity groups in a one-way AI-consultation mode that limits collaborative depth. Low reciprocity may reflect cognitive offloading to GenAI during complex tasks rather than the active cultivation of reciprocal ties, echoing concerns about overdependence noted by Wei et al. (2025).

Figure 4

Structural Characteristics of Role–Role Interaction Networks in Genai-Supported Collaboration



(a) network density by group, (b) degree centralization by group, (c) reciprocity by group, (d) nodes–edges relationship by group, (e) connected components per group, (f) multi-feature network profiles.

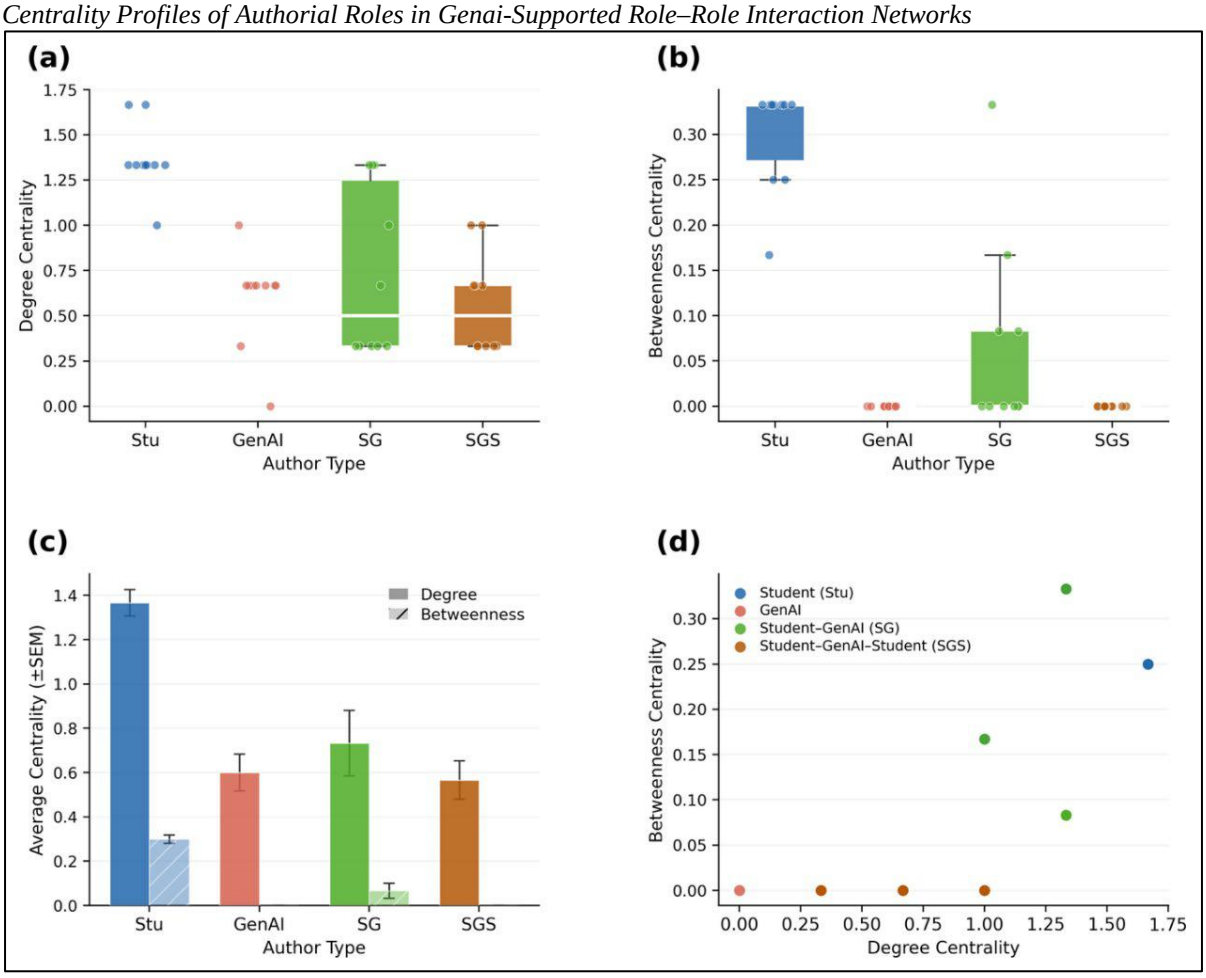
From a nodes-and-edges perspective (Figures 4d and 4f), all groups shared four nodes—Stu, GenAI, SG, and SGS—but the edge counts varied substantially. Groups A and J had seven edges, indicating rich interactions that support decentralized structures and the idea that role diversity can enhance the learning outcomes. Conversely, groups with fewer edges may be constrained by technological adaptation or task design, leading to AI-centered configuration. Multidimensional scaling (MDS) analysis (Figure 3; stress < 0.01) supported the structural reliability; the larger positional variance of the student role suggests within-group dynamism, whereas the relative positions of GenAI and hybrid roles indicate context-dependent interaction patterns. For instance, in high-density groups, AI roles are more embedded within human networks, facilitating a shift from AI-centered to distributed patterns (Feng 2025).

RQ3: Structural functions and centrality profiles of authorial roles

Students functioned as the primary network hub, with the highest degree (1.367 ± 0.189) and betweenness (0.300 ± 0.058) centrality, whereas GenAI and the system-generated scaffold occupied peripheral positions (betweenness = 0), confined to content generation and record-keeping. The student-generated scaffold showed moderate degree centrality (0.733 ± 0.466) but negligible betweenness (0.067 ± 0.110), indicating limited bridging capacity. This pattern underscores the difficulty of positioning GenAI or its outputs as knowledge integrators in the absence of targeted instructional scaffolds (Figures 5a, 5b, 5d).

This asymmetry (Figure 5c) indicates that students shoulder the primary bridging and coordination burden in knowledge-building discourse, consistent with findings that human agents occupy network hubs in hybrid teams (Feng, 2025) and that GenAI's low betweenness reflects inadequate scaffolding rather than inherent limitations (Zhang & Wang, 2025).

Figure 5
Centrality Profiles of Authorial Roles in Genai-Supported Role–Role Interaction Networks



(a) degree centrality by authorial role, (b) betweenness centrality by authorial role, (c) relationship between degree and betweenness centrality, (d) mean centrality by authorial role.

Contrary to our hypothesis, SGS did not attain the expected brokerage position; the betweenness difference was −0.122, diverging from prior expectations (Feng, 2025). This indicates that under current scaffolds, deeper human–AI integration requires further optimization to elevate network brokerage. This finding underscores the complexity of GenAI involvement, potentially constrained by students' critique-and-integration skills or by task context, echoing Sun and Huang's (2025) assertion of the need for hybrid frameworks to optimize role allocation. Overall, these centrality patterns suggest that educational practice should focus on strengthening the brokerage functions of SGS and SG through iterative scaffold design to achieve a more balanced knowledge flow.

Conclusion

This study analyzed 527 discussion notes from 54 undergraduates using authorial role classification and social network analysis to examine how GenAI participates in collaborative discourse within a computational thinking course. Results demonstrate that students function as primary network hubs while GenAI and system-generated scaffolds occupy peripheral positions with zero bridging capacity, revealing a critical gap: current scaffolds fail to position GenAI as a knowledge mediator. This finding underscores the necessity of refined prompt design and instructional scaffolding strategies to transform GenAI from a content provider into a collaborative partner that supports peer knowledge integration. Limitations include classifier generalizability, reliance on discussion notes without learner interviews, and static network analysis that cannot capture temporal dynamics. Future research should examine learner agency in GenAI-supported environments, cross-cultural applicability, and evolving interaction patterns using longitudinal and mixed-method approaches.

 © ISLS

References

Bereiter, C., & Scardamalia, M. (2014). Knowledge building and knowledge creation: One concept, two hills to climb. In S. C. Tan, H. J. So, & J. Yeo (Eds), *Knowledge Creation in Education* (pp. 35–52). Springer Singapore. https://link.springer.com/10.1007/978-981-287-047-6_3

Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education, 16*(12), 2165–2185. https://doi.org/10.1186/s41239-023-00411-8

Chen, B., Zhu, X., & Díaz del Castillo H., F. (2023). Integrating generative AI in knowledge building. *Computers and Education: Artificial Intelligence, 5*, 100184. https://doi.org/10.1016/j.caeai.2023.100184

Cress, U., & Kimmerle, J. (2023). Co-constructing knowledge with generative AI tools: Reflections from a CSCL perspective. *International Journal of Computer-Supported Collaborative Learning, 18*(4), 607–614. https://doi.org/10.1007/s11412-023-09409-w

Feng, S. (2025). Group interaction patterns in generative AI‐supported collaborative problem solving: Network analysis of the interactions among students and a GAI chatbot. *British Journal of Educational Technology, 56*(5), 2125‐2145. https://doi.org/10.1111/bjet.13611

Gong, R., Jiang, R., Guo, C., Hu, W., & Li, Y. (2024). Roles emerging during the knowledge construction process in collaborative learning: Does a generative AI-support chatbot matter? *Proceedings of the 2024 16th International Conference on Education Technology and Computers*, 8–16. https://doi.org/10.1145/3702163.3702165

Grunspan, D. Z., Wiggins, B. L., & Goodreau, S. M. (2014). Understanding Classrooms through Social Network Analysis: A Primer for Social Network Analysis in Education Research. *CBE—Life Sciences Education, 13*(2), 167–178. https://doi.org/10.1187/cbe.13-08-0162

Lee, A. V. Y., Tan, S. C., & Teo, C. L. (2023). Designs and practices using generative AI for sustainable student discourse and knowledge creation. *Smart Learning Environments, 10*(1), 59. https://doi.org/10.1186/s40561-023-00279-1

Ruiz-Rojas, L. I., Salvador-Ullauri, L., & Acosta-Vargas, P. (2024). Collaborative Working and Critical Thinking: Adoption of Generative Artificial Intelligence Tools in Higher Education. *Sustainability, 16*(13), 5367. https://doi.org/10.3390/su16135367

Scardamalia M., & Bereiter C. (2006). Knowledge building: Theory, pedagogy, and technology. In Sawyer R. K. (Ed.), *Cambridge Handbook of the Learning Sciences* (pp. 97–118). Cambridge University Press. https://www.cambridge.org/core/books/cambridge-handbook-of-the-learning-sciences/knowledge-building/57DC9E9EABD2CFC9BCBE395CDA8E375D

Sun, Y., & Huang, Y. (2025). Exploring the Role of Generative AI in Collaborative Lesson Planning for Pre‐Service Teachers. *Journal of Computer Assisted Learning, 41*(4). https://doi.org/10.1111/jcal.70098

Wang, X., Liu, Q., Pang, H., Tan, S. C., Lei, J., Wallace, M. P., & Li, L. (2023). What matters in AI-supported learning: A study of human-AI interactions in language learning using cluster analysis and epistemic network analysis. *Computers & Education, 194*, 104703. https://doi.org/10.1016/j.compedu.2022.104703

Wei, X., Wang, L., Lee, L.-K., & Liu, R. (2025). The effects of generative AI on collaborative problem-solving and team creativity performance in digital story creation: An experimental study. *International Journal of Educational Technology in Higher Education, 22*(1), 23. https://doi.org/10.1186/s41239-025-00526-0

Yin, R. K. (1981). The Case Study Crisis: Some Answers. *Administrative Science Quarterly, 26*(1), 58. https://doi.org/10.2307/2392599

Zhai, X. (2024). Transforming Teachers' Roles and Agencies in the Era of Generative AI: Perceptions, Acceptance, Knowledge, and Practices. *Journal of Science Education and Technology*. https://doi.org/10.1007/s10956-024-10174-0

Zhang, H., & Wang, Q. (2025). How could GenAI work on in-service teachers' knowledge building process? An empirical study based on epistemic network analysis. *International Journal of Educational Technology in Higher Education, 22*(1), 47. https://doi.org/10.1186/s41239-025-00544-y

Acknowledgments

This research was supported by the Guangdong Provincial Education Science Planning Project (No. 2023GXJK266) and the Guangdong Provincial Philosophy and Social Sciences Planning Project (No. GD24CJY33). The authors would like to thank the reviewers for their valuable comments and constructive suggestions, which greatly helped improve the quality of this manuscript.

International Society of the Learning Sciences

Towards Teachers-AI Collaboration for Professional Development in Learning Design

Shatha N. Alkhasawneh, Davinia Hernández-Leo
shatha.alkhasawneh@upf.edu, davinia.hernandez-leo@upf.edu
Universitat Pompeu Fabra, Barcelona, Spain

Abstract: This study examines how Generative Artificial Intelligence (GenAI) can enhance learning design (LD) by exploring stakeholders' preferences for five conceptual GenAI-based conversational agent functions—*Authoring Assistant, Inspiration Hub, Design Feedback, Possibility Engine*, and *Ethics Guardian*—adapted from the *GenAI Speculative Functions Framework for LD*. Proposed for the TRAILS teacher professional development platform, these functions target underexplored early phases of LD. Using a convergent mixed-methods design, 56 participants evaluated these functions, their support needs across design phases, and related opportunities and concerns. Findings reveal a two-pillar demand: creative inspiration, prioritized in the *Investigation* phase through the *Inspiration Hub*, and drafting efficiency, prioritized in the *Authoring* phase through the *Authoring Assistant*. The analysis revealed role-dependent value: experienced practitioners favored generative and creativity-oriented functions, while prospective teachers emphasized diagnostic and feedback support. The *Ethics Guardian* emerged as an essential ambient safeguard. The study offers practical guidance for integrating human-centered GenAI into teacher professional development platforms and human-AI collaboration in LD.

Introduction and pedagogical background

The rapid development of Generative Artificial Intelligence (GenAI) offers both opportunities and challenges for teachers. Rather than replacing teachers, GenAI serves as a collaborative tool that enhances teaching practice by providing evidence, creative ideas, and refined materials for lesson planning, activity design, assessment, content generation, and timely feedback (Bond et al., 2024; Giannakos et al., 2024). This perspective of GenAI as a productive partner is strongly reinforced by conceptual frameworks such as *Human-Centered AI* (Shneiderman, 2020), *Human–AI Collaboration* (Akata et al., 2020), and *Hybrid Intelligence* (Holstein et al., 2020). These perspectives emphasize GenAI's potential to enhance creativity, inclusivity, and efficiency while keeping human judgment central. Sharples (2023) suggests that GenAI can take on social roles in learning—such as *collaboration coach, co-designer, or possibility engine*—shifting from simple assistance to more dialogic and exploratory interaction. However, constructive engagement of GenAI requires teachers to balance its benefits with awareness of its limitations and their responsibility as critical thinkers and ethical practitioners (Divjak et al., 2025; Giannakos et al., 2024). Key concerns include over-reliance, reduced critical thinking, bias, and broader pedagogical and ethical risks. Scholars caution that premature or uncritical adoption could hinder rather than advance education. Therefore, effective GenAI integration requires teacher professional development that offers structured support for both tool use and pedagogical design (Giannakos et al., 2024; Hernández-Leo et al., 2025).

Despite its recognized value, Learning Design (LD) practices remains underused in both schools and higher education, due to barriers such as limited time, low confidence with design methods, insufficient digital skills, and lack of institutional and professional development support (Dagnino et al., 2018). A line of effort to increase LD adoption is integrating supportive tools within dedicated professional development environments (Beardsley et al., 2023). The rise of GenAI intensifies this challenge while also opening new possibilities (Alkhasawneh, Lampropoulos, et al., 2025). Consequently, there is a growing need for tools that can both scaffold teachers' design practices and promote the responsible and ethical use of GenAI in education (Albó et al., 2024; Ravi et al., 2025; Yan et al., 2025). In this context, our study addresses this wider problem by using one particular professional development platform as a methodological case to engage stakeholders in concrete reflection and prioritization of GenAI functions for teacher professional development. The insights gained may inform and extend to other LD and professional development contexts. To address these combined needs, the TRAILS teacher professional development platform was developed within the Erasmus+ project "Towards the Responsible and Constructive Use of Generative AI in Language Learning at School." TRAILS functions as a professional development community, offering a collaborative environment for teachers to co-design GenAI-enhanced learning activities with a strong emphasis on pedagogical quality and responsible adoption. Its core novelty lies in integrating ethical reflection directly into the design workflow. Unlike tools that focus primarily on GenAI's functional capabilities (e.g., content generation), the platform guides teachers to critically address ethical issues

 © ISLS

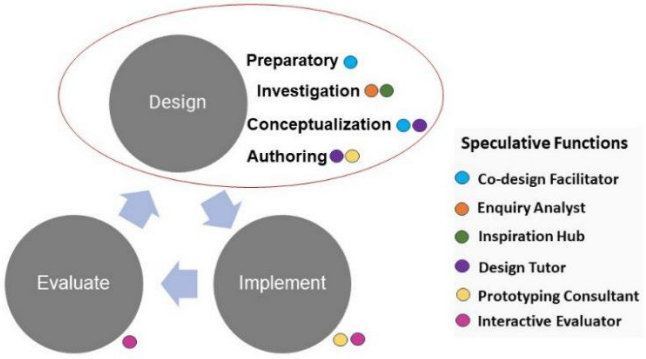
such as plagiarism, bias, and authorship when positioning GenAI as a learning partner in the design of learning activities. This approach promotes authentic student engagement and may contribute to improved learning outcomes (Giannakos et al., 2024). The platform combines collaborative lesson design with inquiry-driven teacher reflection and offers a shared digital space for documenting designs and exchanging resources (Saar et al., 2026).

Despite the platform's practical orientation toward ethics and community, research on how the platform can offer GenAI features to support teacher-AI collaboration when exploring and using the platform towards developing their learning design skills remains in its early stages, with clear calls for empirical studies to validate proposed functions and evaluate their relevance for teachers' collective practices (Hernández-Leo et al., 2025). To structure GenAI support within the platform and address this gap, we drew on the *GenAI Speculative Functions Framework for Learning Design* (see Figure 1; Giannakos et al., 2024). This framework is central to our approach as it positions GenAI as a collaborative design partner—not a replacement for teacher expertise—and is explicitly intended to complement and extend human creativity and professional judgment.

Building on this, we situate the framework within the established structure of the *LD lifecycle*. This design process is commonly organized into four interconnected phases: *Preparation*, *Investigation*, *Conceptualization*, and *Authoring*. In the *Preparation* phase, teachers define the context and learning objectives; in the *Investigation* phase, they analyze learners' needs and contextual constraints; the *Conceptualization* phase involves structuring pedagogical ideas into coherent strategies; and the *Authoring* phase translates these strategies into concrete instructional materials (Giannakos et al., 2024). While all phases are important, the early stages—*Preparation, Investigation,* and *Conceptualization*—remain comparatively underexplored, despite their critical importance for pedagogical coherence and contextual relevance (Ch'ng, 2023; Parsons & Curry, 2024). The *GenAI Speculative Functions Framework* addresses this gap by outlining potential roles for GenAI to support teachers in these crucial early phases through scaffolding processes such as brainstorming, inquiry, structuring, and refinement. Specifically, the framework introduces five GenAI speculative functions, each conceived to complement the teacher's expertise (Alkhasawneh, Hernández-Leo, et al., 2025; Alkhasawneh, Lampropoulos, et al., 2025; Giannakos et al., 2024). These functions include the *Co-Design Facilitator*, which supports collaboration among stakeholders; the *Enquiry Analyst*, which generates insights into learner needs and contextual challenges; the *Inspiration Hub*, which sparks creativity by drawing on prior designs and community examples; the *Design Tutor*, which provides pedagogically grounded feedback; and the *Prototyping Consultant*, which assists in refining and formatting materials for implementation. Figure 1 below illustrates how these functions align with the key phases of the LD process.

Figure 1

GenAI Functions in the LD Lifecycle (adapted from Giannakos et al., 2024).



Drawing on this theoretical foundation, we adapted this framework to the platform's context and proposed five candidate *conversational agent functions* conceptually designed to make the speculative roles actionable for teachers. These functions are: the *Authoring Assistant*, which helps draft or refine task descriptions, learning objectives, and activity instructions; the *Inspiration Hub*, which suggests ideas, activity formats, or real-world examples by leveraging existing learning designs on the platform; the *Design Feedback* function, which reviews design drafts to strengthen their alignment, clarity, and engagement; the *Possibility Engine*, which proposes alternative activity structures or creative variations of learning activities; and finally, the *Ethics Guardian*, which monitors responsible GenAI use and addresses teacher concerns about originality and plagiarism.

While some *GenAI speculative functions* were directly adapted into corresponding *conversational agent functions*—for instance, the *Inspiration Hub* was retained and the *Design Tutor* evolved into *Design Feedback*—others were reinterpreted to better align with the specific pedagogical and ethical orientation of the platform. This context-sensitive adaptation was informed by the platform's specific requirements and driven by the authors'

professional judgment, prioritizing alignment with collaborative design practices and reinforced safeguards for responsible GenAI use. Accordingly, the *Co-Design Facilitator* evolved into the *Possibility Engine*, with the latter focusing on generating alternative design options rather than merely simulating role-play. The *Enquiry Analyst* was fundamentally reimagined as the *Ethics Guardian*, shifting focus from learner profiling—which raised privacy concerns—to safeguarding responsible GenAI use by addressing issues of originality, plagiarism, and integrity. Finally, the *Prototyping Consultant* was adapted into the *Authoring Assistant*, emphasizing the drafting and refinement of instructional elements over technical formatting (see Table 1 below).

Table 1
Mapping GenAI Speculative Functions to Conversational Agent Functions.

GenAI Speculative Function	Conversational Agent Function	Adaptation Rationale
Co-Design Facilitator	Possibility Engine	From simulating collaboration to offering alternative activity structures and creative options.
Enquiry Analyst	Ethics Guardian	From learner profiling to safeguarding responsible GenAI use (e.g., originality, plagiarism, and integrity).
Inspiration Hub	Inspiration Hub	Retained to suggest ideas, activity formats, and real-world examples sourced from the platform's repository.
Design Tutor	Design Feedback	From pedagogical tutoring to feedback on alignment, clarity, and engagement.
Prototyping Consultant	Authoring Assistant	From technical refinement to drafting and refining objectives, tasks, and instructions.

Given these adaptations, a practitioner-oriented validation was required to assess the relevance and perceived value of the proposed *conversational agent functions* for collaborative educational practice. This is especially important in emerging fields where empirical data is scarce, making practitioner evaluations indispensable for generating critical insights that strengthen research validity and applicability (Zondervan-Zwijnenburg et al., 2017). Building on the adapted *GenAI Speculative Functions Framework*, this paper presents a conceptual validation study. It investigates how five candidate *conversational agent functions—Authoring Assistant, Inspiration Hub, Design Feedback, Possibility Engine, and Ethics Guardian*—can be embedded within the authentic LD environment of the TRAILS platform (used here as a case study). This process focused on assessing the conceptual relevance and perceived value of these proposed functions, rather than their fully implemented performance. Using aggregated data from 56 participants (teachers, experts, and researchers) collected between April and July 2025, the study explores: participants' preferences and rationales for the candidate functions; their needs across LD phases; the validation of speculative-function rankings across participant groups; and the opportunities and concerns for GenAI integration in LD. These findings inform phased, responsible GenAI integration into TRAILS and contribute to the wider discourse on evidence-informed, ethical, and collaborative uses of GenAI in LD. The study is guided by the following research questions (RQs):

- **RQ1.** How do teachers, experts and researchers perceive and prioritize candidate *conversational agent functions* aligned with the *GenAI Speculative Functions Framework for LD*, and what rationales underpin their preferences?
- **RQ2.** In which phases of the design process do teachers, experts and researchers perceive the most urgent need for GenAI-supported assistance?
- **RQ3.** How do *GenAI Speculative Function* rankings converge or diverge across practicing teachers, researchers and experts versus prospective teachers?
- **RQ4.** What perceived opportunities and risks do teachers, experts and researchers identify for responsible GenAI integration in the platform, and what design and policy implications follow?

Methodology

Study design, participants, and procedure

This study employed a convergent mixed-methods design (Creswell & Creswell, 2017), combining quantitative and qualitative data to gain a comprehensive understanding of teachers', experts', and researchers' perspectives on GenAI support for LD. The study pursued four main aims: (a) to identify teachers' preferences and priorities for a GenAI assistant—operationalized as a GenAI-based *conversational agent* aligned with the *GenAI Speculative Functions Framework*; (b) to determine in which phases of the LD process (*Preparation, Investigation, Conceptualization*, and *Authoring*) participants perceive the greatest need for GenAI-supported assistance; (c) to validate the *GenAI Speculative Functions Framework* by ranking the functions according to their

perceived value, and examining how these rankings relate to participants' *conversational agent functions* preferences; and (d) to identify the perceived opportunities and risks associated with GenAI integration in the platform, which would then be used to generate evidence-informed design and policy implications for the platform's development. A total of 56 participants took part between April and July 2025. Participants were organized into four distinct cohorts representing diverse professional perspectives: (1) 20 experts in GenAI, LD, and Educational Technology (EdTech); (2) 13 higher education teachers and researchers; (3) 10 prospective teachers enrolled in a Master's program; and (4) 13 secondary school teachers. This cohort structure ensured small, interactive groups and enabled open discussions grounded in shared professional contexts while enriching the study with multidisciplinary and cross-sectoral insights. Each cohort participated in a single workshop facilitated by the same expert to maintain consistency in delivery. All workshops began with an orientation to the TRAILS platform to ensure familiarity with its features before engaging in a design task. Participants were subsequently asked to design use cases for GenAI-supported learning activities within the TRAILS platform, aligned with their teaching and institutional contexts.

Afterward, participants were introduced to the *GenAI Speculative Functions Framework for LD*, mapped to five candidate conversational agent functions and their conceptual roles within the TRAILS platform (*Authoring Assistant, Inspiration Hub, Design Feedback, Possibility Engine,* and *Ethics Guardian*), as well as the four phases of the LD process (*Preparation, Investigation, Conceptualization,* and *Authoring*). Participants then completed the following tasks: (1) A 10-point Likert scale rating their current use of GenAI/LLMs for LD. (2) A *conversational agent function* preferences task (multiple selection; indicating which functions participants considered useful) and a prioritization task (single choice; identifying the one function seen as most important), accompanied by short written justifications for their selections, allowing us to assess the perceived conceptual value and rationale underpinning their choices. (3) A phase-needs task (selecting up to two LD phases where GenAI support was most needed). (4) A ranking of the five *GenAI Speculative Functions* on a scale from 1 to 5, where 1 represented the highest perceived value and 5 the lowest. This specific task was completed individually by two cohorts only and as a group consensus activity for one cohort, in which participants discussed and agreed on a shared ranking. (5) Open-ended reflections described participants' perceived opportunities and concerns regarding the responsible integration of GenAI in LD, which informed subsequent design and policy implications. The cohort representing secondary school teachers did not participate in the ranking task of the *GenAI Speculative Functions* due to time constraints during the workshops, resulting in a smaller sample size for that component. The exact *N* is reported for each analysis. Participation was voluntary, informed consent was obtained, and all responses were pseudonymized in accordance with GDPR-aligned institutional ethics protocols.

Data analysis

Quantitative data were analyzed using descriptive statistics (frequencies, means, medians, modes, and standard deviations) to summarize patterns in *conversational agent function* preferences, prioritization, and perceived needs across the design phases. Given the ordinal nature of the measures and the relatively small sample size, descriptive summaries were the most appropriate analytical approach. For the *GenAI Speculative Functions* validation, results from the experts, teachers, and researchers were aggregated and analyzed based on their individual rankings, while the prospective teachers' group-based ranking was analyzed using *Kendall's coefficient of concordance (W)* to measure the level of consensus. Qualitative data were analyzed using *thematic analysis* (Braun & Clarke, 2006), combining *deductive coding*—guided by the *GenAI Speculative Functions Framework* and the phases of the design process—with *inductive coding* to capture emergent themes. Coding was conducted by one trained researcher and peer-checked by the co-author to ensure reliability and fidelity to participants' perspectives. Triangulation across instruments (function preferences and prioritization, phase needs, and written justifications) further enhanced validity. Finally, quantitative and qualitative findings were integrated to link observable trends with participants' rationales, providing a cohesive interpretation of how GenAI can enhance creativity, efficiency, and pedagogical decision-making in LD.

Results

Participants' context and GenAI use

Participants reported varied levels of prior engagement with GenAI and LLMs in their LD practices. On a 10 -point Likert scale (where 1 represents "never" and 10 represents "very frequently"), usage clustered around the midpoint (mean = 5.3; median = 5; mode = 5), with a standard deviation of 2.8 reflecting substantial variability across participants. In total, 43% reported low use (scores of 2–4), 29% moderate use (5–7), and 27% high use (8–10). This distribution illustrates a continuum ranging from minimal reliance and sporadic experimentation to systematic and deeply integrated use. Qualitative accounts provided further insight into these patterns. Those with

low engagement tended to use GenAI sporadically, often for minor modifications, rephrasing, or occasional inspiration. Participants with moderate engagement described employing it more deliberately—for example, to co-design activities, generate assessments, explore creative alternatives, or refine instructional materials. Those with high engagement portrayed GenAI as a daily companion, serving as both a productivity enhancer and a creative partner in lesson planning, assessment, and pedagogical innovation. Collectively, these findings suggest a continuum of adoption, with over half of participants (56%) reporting moderate to high reliance, highlighting an evolving landscape in which GenAI is steadily becoming embedded in LD practices.

Participants' preferences for conversational agent functions, prioritizations, and justifications

Across the five proposed *conversational agent functions*, the *Inspiration Hub* was the most frequently selected (33/56; 58%), followed by *Design Feedback* and *Possibility Engine* (24/56; 42% each), with the *Authoring Assistant* and *Ethics Guardian* close behind (22/56; 39% each). When participants were asked not only to indicate multiple preferences but also to identify a single highest priority (38 respondents completed this task), the *Inspiration Hub* again led (17 first-place votes). The *Authoring Assistant*—despite fewer overall selections—was ranked first by half of those who selected it (11/22). *Design Feedback* (5 first-place votes) and *the Possibility Engine* (2) were seen as valuable but secondary enhancers, while the *Ethics Guardian* (3) was viewed as an important safeguard rather than a core interactive feature. These contrasts between overall preferences and top-ranked priorities are illustrated in Table 2 and Figure 2 below, which compares the relative preferences and prioritization of each function, along with the key perceived roles participants associated with them.

Table 2

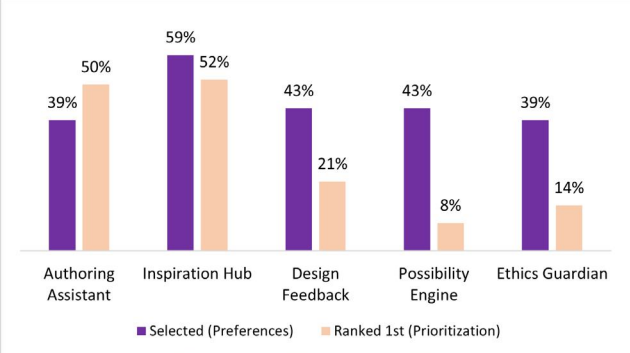
Participants' Preferences and Prioritization of Conversational Agent Functions (N=56; only 38 participants completed the prioritization task).

Conversational Agent Functions	Preferences	Prioritization	Preferences %	Prioritization%	Main Perceived Role
Inspiration Hub	33	17	59%	52%	Core creative partner
Authoring Assistant	22	11	39%	50%	Drafting & structuring
Design Feedback	24	5	43%	21%	Quality refinement
Ethics Guardian	22	3	39%	14%	Responsible safeguard
Possibility Engine	24	2	43%	8%	Alternative variations

The qualitative analysis of participants' justifications revealed three inductively derived themes. (1) Creativity and efficiency—participants emphasized that the *Inspiration Hub* helped overcome design blocks and surface diverse activity formats and examples, while the *Possibility Engine* was valued for offering viable variations without interrupting workflow. (2) Drafting and clarity—the *Authoring Assistant* was praised, especially by practicing and prospective teachers, for accelerating the writing and refinement of objectives, task descriptions, and instructions; the *Design Feedback* function was perceived as offering a constructive "second opinion", helping participants refine alignment, clarity, and engagement in their designs. (3) Responsible use—the *Ethics Guardian* was consistently described as an essential safeguard operating in the background (e.g., plagiarism/originality checks and transparent flagging with explanations), expected to be ever-present yet unobtrusive. Participants also called for personalization to goals, learner level, and subject, and preferred on-demand help over intrusive prompts.

Figure 2

Conversational Agents' Functions Preferences vs. Prioritization.



 © ISLS

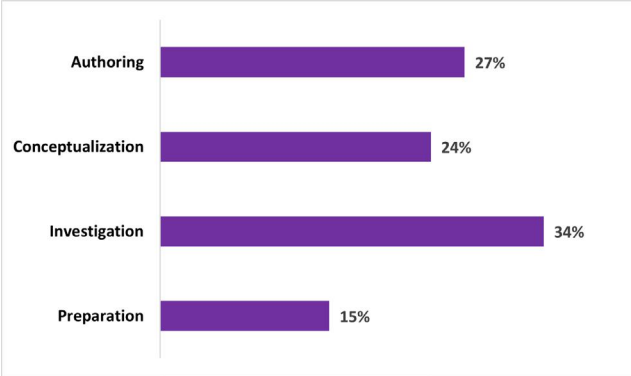
Together, these patterns indicate a two-pillar design demand: (1) A creative catalyst (*Inspiration Hub*) to broaden pedagogical options early, and (2) A drafting scaffold (*Authoring Assistant*) to translate ideas into high-quality materials efficiently. Feedback and variation tools (*Design Feedback, Possibility Engine*) are valued complements, while ethics features are expected as ambient safeguards rather than foregrounded interactions.

Participants' demand for GenAI support by design phase

Across the four design phases, participants (N = 56) most often requested GenAI support via a *conversational agent* during the *Investigation* phase (34%), citing support for exploring pedagogical strategies, locating resources, and interpreting learner needs. The next highest need was the *Authoring* phase (27%), where teachers sought assistance in drafting and refining materials to translate ideas into implementable activities. The *Conceptualization* phase (24%) followed, reflecting interest in structuring ideas and sequencing activities to overcome creative blocks. The *Preparation* phase (15%) was least selected, suggesting that defining goals and highlighting context remains largely grounded in teachers' professional judgment.

Figure 3

Participants' Preferences for GenAI Support by Design Phase.



Justifications revealed that the *Investigation* phase is time-consuming and cognitively demanding, making GenAI support for resource exploration, strategy scanning, and needs analysis particularly valuable. The *Authoring* phase was described as tedious yet highly amenable to GenAI-enabled drafting and refinement. The *Conceptualization* phase was framed as one in which GenAI can act as a co-designer, helping to structure and sequence ideas. By contrast, the *Preparation* phase was viewed as best led by teachers' expertise, as it requires goal-setting, curricular alignment, and ethical and contextual awareness beyond GenAI's current capacity—positioning GenAI a as minor, supplementary aid at this phase. Overall, the results indicate a clear hierarchy—from *Investigation* to *Authoring* to *Conceptualization* to *Preparation*—that can guide phased integration: prioritizing research/inspiration and drafting supports, positioning creative structuring as a complementary aid, and treating *Preparation* as predominantly human-led. These *phase-based* findings reinforce the patterns observed in the *conversational agent function* preferences and prioritization analysis. The strong demand for support in the *Investigation* and *Authoring* phases echoes the prioritization of the *Inspiration Hub* and *Authoring Assistant* as core functions, while *Conceptualization* aligns with complementary creativity and structuring aids such as the *Possibility Engine*. *Preparation*, by contrast, remains grounded in professional expertise, resonating with the view that pedagogical and ethical framing should be safeguarded but not foregrounded. Taken together, the convergence of *function-level* and *phase-level* analyses emphasizes a two-pillar design need for the platform: (1) GenAI-enabled inspiration and exploration in the early stages of design, and (2) GenAI-enhanced drafting and refinement in the later stages, with other functions serving as valuable complements rather than primary drivers.

Participants' validation of GenAI speculative functions

The validation of the *GenAI Speculative Functions* was conducted with three different cohorts who applied different ranking methods according to their context and composition. The first two—experts in GenAI, LD, and EdTech (n = 20) and higher education teachers and researchers (n = 13)—jointly formed a combined group that completed individual rankings of the five functions on a scale from 1 (highest perceived value) to 5 (lowest). The third group—prospective teachers enrolled in a Master's program (n = 10)—conducted a group-based ranking, where consensus levels were calculated using *Kendall's coefficient of concordance (W).*

Among the combined group of experts, teachers, and researchers, the *Inspiration Hub* emerged as the top-ranked function (mean = 1.67; median = 1; mode = 1), indicating strong consensus on its importance as a

creative catalyst supporting idea generation and conceptual exploration. The *Design Tutor* (mean = 2.00; median = 2) and *Co-Design Facilitator* (mean = 2.06; median = 2) followed closely, valued for their formative and collaborative potential. The *Prototyping Consultant* (mean = 2.58; median = 2) and *Enquiry Analyst* (mean = 2.61; median = 3) were rated moderately, reflecting that while participants recognized their usefulness for refinement and contextual analysis, they were viewed as less immediately essential. The relatively low standard deviations across functions (ranging from 0.94 to 1.56) indicate moderate agreement within this group, particularly regarding the consistent prioritization of creativity- and feedback-oriented functions. The prospective teachers, in contrast, reached an almost unanimous consensus (*Kendall's W = 0.96*) in their group-based rankings. They placed the *Design Tutor* first across all groups, followed by the *Enquiry Analyst*. The *Inspiration Hub* was consistently ranked in the middle (third or fourth), while the *Prototyping Consultant* and especially the *Co-Design Facilitator* were ranked as least relevant (fourth and fifth, respectively). This pattern suggests that early-career teachers prioritize diagnostic and feedback-oriented tools that support the refinement and evaluation of designs over generative or collaborative functions.

Taken together, these findings reveal both convergence and role-specific differentiation in how participants perceive the value of *GenAI Speculative Functions for LD*. Across groups, there was clear agreement on the importance of creativity and feedback as core areas of GenAI support. The combined group of experts, teachers, and researchers prioritized the *Inspiration Hub* and *Design Tutor*, emphasizing GenAI's capacity to act as a creative engine for sparking ideas, facilitating the exploration of key concepts, and strengthens pedagogical alignment. Their appreciation of the *Co-Design Facilitator* further reflected recognition of GenAI's potential to scaffold collaborative reflection and joint problem-solving in professional design settings. In contrast, the prospective teachers placed the greatest value on the *Design Tutor* and *Enquiry Analyst*, revealing a focus on diagnostic and feedback-oriented tools that provide structure, formative guidance, and support for decision-making during early-career practice. This progression—from inspiration and collaboration among experts to feedback and evaluation among novice teachers—suggests that GenAI functions resonate differently depending on users' experience and design autonomy. Accordingly, a phased integration approach within the platform—beginning with inspiration and feedback features and progressively expanding to collaborative and prototyping tools—would best align with the evolving needs and capacities of diverse teacher groups.

Opportunities and concerns for GenAI integration in LD

Thematic coding of open responses revealed a balance of optimism and caution. On the opportunity side, participants highlighted practical, high-leverage uses of GenAI in LD. The most frequently cited opportunities were: (a) Creativity support—rapid idea generation, alternative activity formats, and examples to overcome design blocks; (b) Efficiency—time savings in drafting and refining objectives, task descriptions, resources, and variants for different levels; (c) Differentiation—tailoring and adapting materials to learner profiles; (d) Prototyping/simulation—quickly producing illustrative artefacts for lessons; and (e) Pedagogical breadth—surfacing a wider repertoire of pedagogical approaches. Collectively, these point to GenAI's value as both a creative catalyst and a productivity amplifier in teachers' everyday design work. Counterbalancing these opportunities, concerns clustered around: (a) Professional risks—over-reliance that may hinder teacher creativity, reduce critical reflection, or homogenize outputs; (b) Quality risks—inaccurate, misleading, or decontextualized content requiring careful review; (c) Ethics and integrity—plagiarism, opaque sourcing, and bias; (d) Data and privacy—uncertainty about using learner data to personalize suggestions; and (e) Practical barriers—limited time and training to use tools effectively and integrate them into real classroom constraints. Several participants stressed that GenAI should assist, not replace, pedagogical judgment.

Design and policy implications for the platform

Findings point to three requirements for the responsible integration of GenAI into the platform, as perceived by teachers, experts and researchers, to better support and enhance LD practices. First, make creativity and drafting first-class features: participants consistently prioritized the *Inspiration Hub* and *Authoring Assistant*, highlighting the need for seamless flows from ideation to structured drafts. These should form the platform's core interactive functions. Second, build reviewability into all generative support. Teachers emphasized the importance of understanding and critically evaluating GenAI outputs. Embedding *Design Feedback*—providing rationales, alignment checks, and suggested edits—together with verification and sourcing prompts, ensures GenAI acts as a transparent assistant rather than a black box. Third, embed ethics and privacy safeguards by default. An unobtrusive *Ethics Guardian* should flag originality risks with clear explanations, warn against bias or overclaiming, and guide acceptable use (e.g., paraphrasing or attributing). Any use of learner data should be gated by explicit consent and strict minimization practices.

Collectively, these design directions call for a set of guardrails to preserve teacher agency while maximizing the value of GenAI. Human-in-the-loop defaults should require teacher approval or edits before AI-generated content is published or shared. Context controls should allow teachers to specify goals, learner level, and design constraints, ensuring that outputs are tailored rather than generic. Transparency around provenance—showing how and why a suggestion was generated—would build trust, while always-on plagiarism and bias checks could provide teachers with actionable, plain-language explanations and safer alternatives. Finally, capacity building is essential: short, just-in-time tutorials and exemplars should accompany the *Inspiration*, *Authoring*, and *Feedback* features, equipping teachers to use them responsibly and effectively. Given the TRAILS' role in teacher professional development, capacity building needs to be framed within a broader pedagogical strategy that fosters teachers' critical reflection and maintains design autonomy. In this way, TRAILS can balance usability and innovation with accountability and trustworthiness, establishing itself as both a creative catalyst and an ethical safeguard in GenAI-enhanced LD.

Discussion and conclusion

This study set out to investigate how teachers, experts and researchers perceive, prioritize, and rationalize the integration of GenAI into teacher-AI collaboration in LD, as well as the opportunities and concerns they associate with its use. By using the TRAILS Platform as a professional development context and methodological case, we were able to engage stakeholders in concrete reflection and prioritization of *GenAI functions*, directly contributing to the wider discourse on how GenAI can systematically enhance collaborative and reflective LD processes. Building on the *GenAI Speculative Functions Framework for LD* (Giannakos et al., 2024), the study adapted five candidate *conversational agent functions* for conceptual validation. Across analyses, a consistent pattern emerged: GenAI is valued for saving time, enhancing creativity, and scaffolding drafting, while raising concerns about pedagogical depth, professional agency, and ethics.

At the *function level*, the *Inspiration Hub* and *Authoring Assistant* emerged as core tools, reflecting views of GenAI as a creative collaborator that complements human expertise (Sharples, 2023; Holstein et al., 2020). *Design Feedback* and the *Possibility Engine* were seen as enhancers, while the *Ethics Guardian*—though less prioritized—was expected as an embedded safeguard aligned with *Human-Centered AI* principles (Shneiderman, 2020). *Conversational agents* further support the dialogic nature of GenAI, enabling iterative and exploratory interaction (Sharples, 2023). At the *phase level*, demand was strongest in the *Investigation* and *Authoring* phases, echoing literature that identifies these stages as resource-intensive and cognitively demanding (Parsons & Curry, 2024; Ch'ng, 2023). The *Conceptualization* phase played a secondary but useful role for GenAI collaboration, while the *Preparation* phase remained largely teacher-led. *Role-based* differences also emerged: practicing teachers prioritized creativity and inspiration, whereas prospective teachers emphasized feedback and diagnostic support—reflecting how perceptions of GenAI align with professional experience (Dagnino et al., 2018). Overall, findings reflect a balance of optimism and caution. GenAI was valued as a creative and efficiency-enhancing support, particularly for differentiation, but concerns persisted around over-reliance, standardization, ethical risks, and data privacy (Divjak et al., 2025; Bond et al., 2024).

This study has several limitations. The sample size (N = 56) was relatively small, limiting generalizability. The focus on the design phase of the *LD lifecycle* leaves other stages—particularly *Implementation* and *Evaluation*—underexplored. Validation involved only three cohorts, reducing the diversity of perspectives. As a workshop-based study, findings reflect perceived usefulness rather than long-term adoption. Moreover, the study is based on the conceptual validation of candidate *conversational agent functions* that were not yet implemented in the platform, capturing perceived usefulness and prioritization rather than observed usability or impact. Nevertheless, this conceptual validation step remains important for ensuring relevance and aligns with strategies supporting LD adoption (Beardsley et al., 2023).

Building on these insights, future work will focus on empirically evaluating the integrated *conversational agent functions* and their impact on teachers' design practices, efficiency, creativity, pedagogical decision-making, and learner outcomes. In conclusion, this study offers a human-centered prioritization of the *Speculative Functions Framework for LD*. Findings show that teachers value GenAI support that broadens options and improves efficiency, while emphasizing the need for ethical safeguards and preserved professional judgment. A phased rollout centered on *Inspiration* and *Authoring*, supported by transparent feedback and embedded ethics features, addresses these expectations and supports responsible GenAI integration in LD.

References

Akata, Z., Balliet, D., de Rijke, M., Dignum, F., Dignum, V., Eiben, G., Fokkens, A., Grossi, D., Hindriks, K., Hoos, H., Hung, H., Jonker, C., Monz, C., Neerincx, M., Oliehoek, F., Prakken, H., Schlobach, S., van der Gaag, L., van Harmelen, F., … Welling, M. (2020). A Research Agenda for Hybrid Intelligence: Augmenting Human Intellect With

Collaborative, Adaptive, Responsible, and Explainable Artificial Intelligence. Computer, 53(8), 18–28. https://doi.org/10.1109/MC.2020.2996587

Albó, L., Noguera, I., Hernández-Leo, D., Afonso, A., Agostini, D., & Alkhasawneh, S. N. (2024). FLeD Learning Design Tool: Scaffolding Flexible Scenarios Through the Use of Flipped Learning Patterns. In R. Ferreira Mello, N. Rummel, I. Jivet, G. Pishtari, & J. A. Ruipérez Valiente (Eds.), Proceedings of the 19th European Conference on Technology Enhanced Learning (EC-TEL 2024): Technology Enhanced Learning for Inclusive and Equitable Quality Education (Vol. 15160, pp. 209–215). Krems, Austria: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-72312-4_29

Alkhasawneh, S. N., Hernández-Leo, D., & Dimitriadis, Y. (2025). Experts' Views on Generative AI's Speculative Functions in Learning Design. In A. I. Cristea, E. Walker, Y. Lu, O. C. Santos, & S. Isotani (Eds.), Proceedings of the 26th International Conference on Artificial Intelligence in Education (AIED 2025) (pp. 235–242). Palermo, Italy: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-98465-5_30

Alkhasawneh, S. N., Lampropoulos, G., & Hernández-Leo, D. (2025). Generative AI in Learning Design: A Systematic Review. In A. I. Cristea, E. Walker, Y. Lu, O. C. Santos, & S. Isotani (Eds.), Proceedings of the 26th International Conference on Artificial Intelligence in Education (AIED 2025) (pp. 17–30). Palermo, Italy: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-98459-4_2

Beardsley, M., Hernández-Leo, D., & Sánchez-Reina, R. (2023). Using Learning Design Technologies for Teachers' Practice-Driven Research. In Proceedings of ICCE 2023: The 31st International Conference on Computers in Education (pp. 829–897). Matsue, Shimane, Japan: APSCE. https://library.apsce.net/index.php/ICCE/article/view/4730

Bond, M., Khosravi, H., De Laat, M., Bergdahl, N., Negrea, V., Oxley, E., Pham, P., Chong, S. W., & Siemens, G. (2024). A meta systematic review of artificial intelligence in higher education: A call for increased ethics, collaboration, and rigour. International Journal of Educational Technology in Higher Education, 21(1), 4. https://doi.org/10.1186/s41239-023-00436-z

Braun, V., & Clarke, V. (2006). Using Thematic Analysis in Psychology. Qualitative Research in Psychology, 3(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Ch'ng, L. K. (2023). How AI Makes its Mark on Instructional Design. Asian Journal of Distance Education, 18(2), 32-41. https://asianjde.com/ojs/index.php/AsianJDE/article/view/740

Creswell, J. W., & Creswell, J. D. (2018). Research Design: Qualitative, Quantitative, and Mixed Methods Approaches (4th ed.). SAGE Publications.

Dagnino, F. M., Dimitriadis, Y. A., Pozzi, F., Asensio-Pérez, J. I., & Rubia-Avi, B. (2018). Exploring Teachers' Needs and the Existing Barriers to the Adoption of Learning Design Methods and Tools: A Literature Survey. British Journal of Educational Technology, 49(6), 998–1013. https://doi.org/10.1111/bjet.12695

Divjak, B., Svetec, B., Vondra, P., Bađari, J., & Grabar, D. (2025). Learning Design with an AI Assistant. In A. I. Cristea, E. Walker, Y. Lu, O. C. Santos, & S. Isotani (Eds.), Artificial Intelligence in Education (Vol. 15877, pp. 207–220). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-98414-3_15

Giannakos, M., Azevedo, R., Brusilovsky, P., Cukurova, M., Dimitriadis, Y., Hernandez-Leo, D., Järvelä, S., Mavrikis, M., & Rienties, B. (2024a). The Promise and Challenges of Generative AI in Education. Behaviour & Information Technology, 1–27. https://doi.org/10.1080/0144929X.2024.2394886

Hernández-Leo, D., Theophilou, E., Oshima, J., Minematsu, T., Matsueda, K., Naganuma, S., Cress, U., Chen, B., Zhu, X., del Castillo, F. D., Chen, W., Lyu, Q., Chen, X., Zheng, L., & Weinberger, A. (2025). Collaborating with Generative AI for Learning? In Proceedings of the 18th International Conference on Computer-Supported Collaborative Learning-CSCL 2025, pp. 525-533. International Society of the Learning Sciences. https://doi.org/10.22318/cscl2025.882858

Holstein, K., Aleven, V., & Rummel, N. (2020). A Conceptual Framework for Human–AI Hybrid Adaptivity in Education. In I. I. Bittencourt, M. Cukurova, K. Muldner, R. Luckin, & E. Millán (Eds.), Artificial Intelligence in Education (pp. 240–254). Springer International Publishing. https://doi.org/10.1007/978-3-030-52237-7_20

Parsons, B., & Curry, J. H. (2024). Can ChatGPT Pass Graduate-Level Instructional Design Assignments? Potential Implications of Artificial Intelligence in Education and a Call to Action. TechTrends, 68(1), 67–78. https://doi.org/10.1007/s11528-023-00912-3

Ravi, P., Masla, J., Kakoti, G., Lin, G. C., Anderson, E., Taylor, M., & Abelson, H. (2025, April). Co-designing Large Language Model Tools for Project-Based Learning with K12 Teachers. In Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (pp. 1-25). https://dl.acm.org/doi/full/10.1145/3706598.3713971

Saar, M., Ljalikova, A., Theophilou, E., Alkhasawneh, S. N., & Hernández-Leo, D. (2026). Empowering Teachers to Enhance Student High Order Thinking Skills with Artificial Intelligence Educational Tools. In K. Tammets, S. Sosnovsky, R. Ferreira Mello, G. Pishtari, & T. Nazaretsky (Eds.), Proceedings of the 20th European Conference on Technology Enhanced Learning (EC-TEL 2025): Two Decades of TEL – From Lessons Learnt to Challenges Ahead (pp. 442–456). Newcastle & Durham, United Kingdom: Springer Nature Switzerland. https://doi.org/10.1007/978-3-032-03870-8_30

Sharples, M. (2023). Towards Social Generative AI for Education: Theory, Practices and Ethics. Learning: Research and Practice, 9(2), 159–167. https://doi.org/10.1080/23735082.2023.2261131

Shneiderman, B. (2020). Human-Centered Artificial Intelligence: Three Fresh Ideas. AIS Transactions on Human-Computer Interaction, 12(3), 109–124. https://doi.org/10.17705/1thci.00131

Yan, L., Martinez-Maldonado, R., Jin, Y., Echeverria, V., Milesi, M., Fan, J., Zhao, L., Alfredo, R., Li, X., & Gašević, D. (2025). The Effects of Generative AI Agents and Scaffolding on Enhancing Students' Comprehension of Visual Learning Analytics. Computers & Education, 234, 105322. https://doi.org/10.1016/j.compedu.2025.105322

Acknowledgments

This work has been supported by PID2023-146692OB-C33 and 2023-26 CEX2021-001195-M granted by MICIU/AEI/10.13039/501100011033 and KA220-HED-D8B72E6A. D. Hernández-Leo (Serra Húnter) also acknowledges the support of the ICREA Academia programme.

 © ISLS

International Society of the Learning Sciences

Socially Shared Metacognitive Regulation (SSMR): Comparative Analysis of Strategies and Adaptations in Teams with Divergent Outcomes during CPS

Vishwas Badhe, Ramkumar Rajendran
vishwasbadhe@iitb.ac.in, ramkumar.rajendran@iitb.ac.in
Indian Institute of Technology Bombay, Mumbai, India

Abstract: Collaborative problem-solving (CPS) for ill-structured problems is an integral part of various learning settings, where teams apply varied SSMR strategies and perform necessary adaptations. The effective application and adaptation of SSMR strategies contribute to CPS success, making their investigation crucial for understanding SSMR's in CPS. This study undertakes a comparative analysis of SSMR strategy application and adaptation in teams with divergent outcomes during ill-structured CPS. Five engineering teams *($N_{Learners}=20$)* completed four-hour programming-related ill-structured CPS tasks in a closed lab setting. Teams were classified as more and less successful outcome teams (MSOT and LSOT) using an evaluation rubric. Verbal interactions (of 20 hours) were analyzed using a mixed-methods approach combined with comparative thematic analysis. Considerable differences were found between MSOT and LSOT regarding SSMR strategy application and adaptation. These differences are presented through comparative frequency analysis and thematic representations, revealing distinctions in how MSOT and LSOT strategize and adapt during CPS.

Introduction

Collaborative Problem Solving (CPS) is a dynamic process where teams must develop a shared mental model and collective scheme of cognitive interdependence to ensure high-quality participation and shared understanding (Järvelä, 2021; Vogel et al., 2023). This process requires learners to regulate their own cognition, metacognition, and motivation through shared metacognitive monitoring (Järvelä et al., 2013). However, challenges frequently arise from differing prior knowledge and task interpretations, varied task familiarity, and diverse socio-cultural backgrounds (Sobocinski et al., 2021). In such contexts, Socially Shared Metacognitive Regulation (SSMR) becomes critical, as it enables teams to collectively recognize their regulatory needs (Vauras et al., 2021; Seiradakis & Spantidakis, 2019; Iiskala et al., 2011). Socially shared metacognitive regulation (SSMR) involves participants working together in a goal-directed, shared, and reciprocal manner to regulate shared thinking processes during collaborative learning context (Iiskala et al., 2015; Iiskala et al., 2011). SSMR plays an essential role in CPS by guiding teams' cognitive activity through monitoring and control processes to ensure CPS progress toward shared goals (Shin, Lee, & Kim, 2025; Schoor, Narciss, & Körndle, 2015). During ill-structured CPS, teams formulate and apply various SSMR strategies targeting task and content understanding, planning (goal setting), and task performance (Sobocinski et al., 2021). However, due to the paucity of studies investigating SSMR strategies in ill-structured CPS contexts, more empirical evidence is needed to understand their dynamic application (Lobczowski et al., 2021). Moreover, existing studies investigating SSMR strategies have not adequately considered the performance levels of collaborating teams, highlighting the need for research that accounts for both performance outcomes and the dynamic emergence of SSMR across various contexts (Lobczowski et al., 2021).

Beyond strategy application, CPS demands 'on-the-fly' adaptations in SSMR strategies during implementation, as team members negotiate adjustments to cognitive, metacognitive, emotional, motivational, and behavioral states as needed. Such adaptation is considered an efficient form of regulation (Hadwin et al., 2018). Adaptation is described as small-scale when it occurs "on the fly" and aims to optimize the current learning process in response to shared metacognitive monitoring indicating a need to adapt (Hadwin et al., 2018). In contrast, large-scale adaptation refers to transferring regulatory knowledge gained in the current task to future situations (Winne & Hadwin, 1998).

Regarding small-scale adaptations in SSMR, the limited number of existing studies necessitates more empirical evidence across various contexts (Sobocinski et al., 2021). Furthermore, available evidence on small-scale adaptations has predominantly focused on well-performing teams, warranting detailed investigation of SSMR strategies in both more and less successful outcome teams (Sobocinski et al., 2021). Additionally, since existing evidence for small-scale adaptations derives from studies of short duration, investigating these adaptations in longer-duration studies would be valuable (Sobocinski et al., 2019). This study investigates

 © ISLS

SSMR strategy application and small-scale adaptations in divergent outcome teams specifically, more successful outcome teams (MSOT) and less successful outcome teams (LSOT) during ill-structured CPS. We classified teams as MSOT and LSOT based on the premise that SSMR is crucial for collaborative success (Malmberg et al., 2015; Ucan & Webb, 2015) and to explicitly consider the performance factor in our comparison. The ill-structured nature of the tasks increased difficulty, requiring enhanced SSMR strategy application for effective collaboration (Tucker et al., 2020; Iiskala, 2011). Our analysis employed comparative thematic analysis and quantitative measures to examine SSMR strategies and small-scale adaptations between MSOT and LSOT. The research questions (RQs) guiding this study is: 1) how do the SSMR strategies differ between teams with divergent outcomes during ill-structured CPS? and 2) how do the small-scale adaptations in SSMR strategies differ between teams with divergent outcomes during ill-structured CPS?

Methodology

Participants

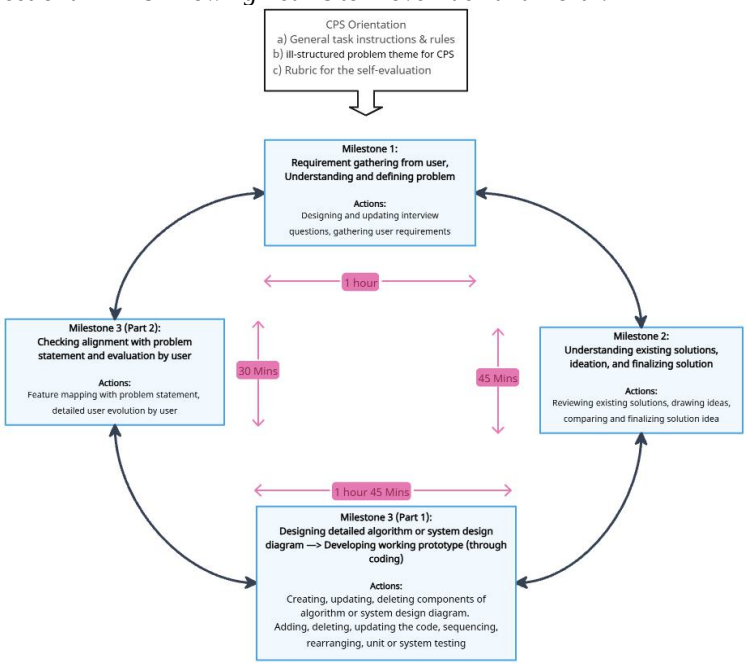
Five teams comprising 20 undergraduate engineering learners (Mean age = 19.2 years, SD = 0.67; 65% Male, 35% Female) from a state university participated in this study with prior informed consent. Participants were recruited from two affiliated engineering colleges and registered voluntarily, forming teams based on familiarity with classmates. Learners were primarily third and fourth-year computer science engineering majors.

CPS Task and study design

The ill-structured problem scenario for learners was: "A 9th-grade student from a vernacular medium school in India struggles with understanding C++ syntax and error messages in English despite having strong mathematical and logical thinking skills. The student needs a programming interface for understanding programming errors in their local language." A dummy user was available throughout CPS for guidance. The CPS required learners to apply software development and programming skills including coding concepts, debugging and troubleshooting (Çoklar and Akçay, 2018). The ill-structured nature allowed teams to choose aligned problem statements, encouraging diverse thought processes, brainstorming, and decision-making. The CPS study design was structured around three major milestones with corresponding tasks, as illustrated in Figure 1. Each milestone contained a heading serving as an objective, indicating tasks and major actions for teams. Separate closed rooms were allocated to each team to ensure an isolated working environment.

Figure 1

The Ill-Structured CPS Process Has Three Major Milestones, Starting with CPS Orientation and Progressing Clockwise, with Bidirectional Links Allowing Teams to Move Back and Forth.



Milestone 1 focused on data gathering and problem analysis tasks. Milestone 2 required reviewing existing solutions and finalizing one idea. Milestone 3 involved developing a low-fidelity prototype and user

evaluation. The four-hour CPS task included comprehensive instructions, tasks, and a self-evaluation rubric (1) used to evaluate team performance. All teams successfully completed the task with varied performance levels.

The study began with a CPS orientation session. Teams were informed this was not a competition and could use any programming language or framework combination but were strictly prohibited from using LLMs, Co-pilot, or AI-based tools, and mobile phones as this could undermine externalizing SSMR strategies during CPS. Participants could consult websites like Google or Stack Overflow for technical details and debugging but not for ready-made solutions. Team members worked in face-to-face, co-located settings using individual laptops with options to share work. A physical whiteboard supported ideation and drawing like system design diagram creation. The facilitator explained the ill-structured problem theme, introduced dummy users, and discussed milestone-wise deliverables, all provided as handouts. The CPS session spanned maximum four hours with soft milestone time allocations as guidelines. Teams retained flexibility to divide responsibilities, work sequentially through milestones 1-3, and revisit previous milestones. Each team submitted deliverables as a single document at session end and demonstrated their low-fidelity prototypes to the dummy user.

Data collection

Data was collected from five teams (N = 20) through videotaped verbal interactions and low-fidelity prototype solutions created using different programming languages. Each member worked on a personal laptop with options to contribute on a shared screen. Verbal interactions were captured using a camera, yielding five team session videos of approximately four hours each, totaling 20 hours. Teams used Google Docs and physical whiteboards to create shared representations such as system design diagrams. Verbal interaction data was primarily analyzed to examine SSMR strategies and adaptation traces, while solution prototypes, Google Docs submissions, and shared representations provided additional contextual clarification when required. Progressive solution prototypes were considered in evaluating teams' performance against the self-evaluation rubric.

Coding scheme and strategy

A three-phase coding process was implemented for data analysis. In the first phase, episodes showing application of SSMR strategies were identified. Verbal interaction containing enactment/application of predefined SSMR strategy by two or more members is considered as an SSMR episode (Isohätälä et al., 2018; Järvenoja et al., 2020). The SSMR strategy episode reflects continued patterns of collaboration and topic, ending with a clear shift in either area. In the second phase, to examine micro-level application of metacognitive skills, SSMR strategy episodes were further analyzed at the statement level using a deductive categorization method of content analysis (Mayring, 2015, p. 95; Ader et al., 2023). A statement, defined as a single conversational turn, represents a thematically consistent verbalization of a metacognitive action or skill by an individual team member (De Backer, 2015). Each statement was coded for the metacognitive skills in SSMR (such as orienting, planning, monitoring, and evaluating) along with corresponding sub-skills (e.g., comprehension monitoring or strategic planning) (Meijer et al., 2006).

In the third phase, to understand and map SSMR strategies applied by team members, statements situated in SSMR strategy episodes were coded using four strategy codes: monitoring/controlling task understanding, monitoring/controlling content understanding, planning task performance, and monitoring/controlling task performance (Lobczowski et al., 2021). Finally, coding was concluded by marking traces of small-scale adaptations (either adaptive or maladaptive). The trace of small-scale adaptation in SSMR strategy was marked as adaptive if the monitoring event revealed the need for regulation and team members used the information gained to strategically control their learning (Winne and Hadwin, 1998), either through making adjustments to enacted strategies or by strategically adapting their task perceptions, plans, and actions (Sobocinski et al., 2021). The trace was marked as maladaptive if the monitoring event revealed the need for regulation but team members failed to use the information to strategically control their learning by making desired adjustments to enacted strategies or failed to adapt their task perceptions, plans, and actions. The target of the small-scale adaptation is marked in the strategy code (Sobocinski et al., 2021). The comprehensive coding scheme used for deductive categorization of statements is shown in Table 1. Two examples (respectively for MSOT and LSOT) of episode identification and statement coding are presented in Table 3a &b.

CPS performance was evaluated by experienced educational technology researchers. The rubric consisted of three milestones, nine objectives and tasks, aligned with 13 deliverables. Each deliverable and process carried a maximum of 10 points, totaling 130 possible points. A 60% cutoff (78 out of 130) was used for classification: teams scoring 78 or higher were categorized as MSOT, and those below as LSOT. Teams 1, 3, and 4 were classified as MSOT with scores of 93, 86, and 98 respectively, while Teams 2 and 5 were classified as LSOT with scores of 74 and 63. Statements in SSMR strategy episodes were coded by trained educational technology researchers.

Table 1
Coding Scheme Used for Deductive Categorization Method While Analyzing SSMR Skill, Strategies, and Adaptation Trace from Verbalized Interactions

	Code	Description	Example
Meta-cognitive regulation skill (1)	Orienting	Team members engage in task analysis, which might result in becoming aware of one's task perceptions or activating one's prior knowledge	*"M3-but there are some problems in react (framework) like routing and all.""*
	Planning	Encompasses selecting and sequencing problem-solving strategies and developing action plans.	*"M2-yeah, just if you store in it, then I will pull it from it (file)"*
	Monitoring	Involves quality control of one's learning or problem-solving, aimed at identifying inconsistencies and at optimizing task execution. It also involves monitoring for comprehension, progress, and collaboration.	*"M1-One second. once it is executing then where its going ? and sending response to where?"*
	Evaluating	Involves learners' self-judgment upon completion of problem-solving. This can be directed at the learning outcomes, the problem-solving process, or the group members' collaboration.	*"M3-don't think other things, focus on what we want to show, we need to integrate everything in next 10-15 mins right?"*
SSMR strategies (2)	M/C task understanding	Monitoring and control processes related to enacting strategies for understanding task including small and final.	*"M4- So yeah, interview questions, through which we would be able to collect more data* *M3- And what types of que we will prepare?"*
	M/C content understanding	Monitoring and control processes related to enacting strategies for understanding content related to their problem.	*"M1- I don't think so it's a (different than problem statement). It is the problem statement I feel."*
	Planning task performance	Processes related to the setting of goals or creation of plans related to task understanding, content understanding, or task performance.	*"M4- Let's ask them the first question. What are the solutions you have tried by yourself? so will come to know.."*
	M/C task performance	Monitoring and control processes related to the tracking of progress or enacting strategies related to task performance.	*"M2- But then this second question will not be useful then? that what type of UI you need."*
Small-scale adaptation trace	Adaptive	Adaptation is an event triggered by monitoring where teams identify regulatory needs and strategically adjust their strategies, perceptions, plans, or actions accordingly.	*"M3- so tell me one thing. You are going to integrate or I should do?* *M2- Shall I integrate?* *M3- Ohk, So my work is dependent on your work. So you do it first."*
	Maladaptive	Maladaptation is the absence of strategic adjustment or implementation of inappropriate responses when monitoring reveals regulatory needs.	*"M4- No, They have given the scenario. So paraphrasing it only, we can make a PS…* *..M4- yeah. Just remove scenario."*

Note: M/C- Monitoring/controlling

The inter-rater reliability was assessed using Cohen's kappa (κ). Of 763 coded statements, 200 (26%) were selected to calculate reliability, which also contained traces of adaptations and maladaptation. Researchers compared results, discussed discrepancies, and resolved them through social moderation. Cohen's kappa values exceeded 0.85 for all constructs (metacognitive skill, type of SSMR strategy, and small-scale adaptations), indicating strong agreement. After establishing reliability, the remaining dataset was independently coded by one researcher

Data analysis

We analyzed 20 hours of videotaped verbal interaction data from five teams using a mixed-methods approach. The analysis employed deductive content analysis combined with comparative thematic analysis (Joffe, 2012) to examine SSMR strategies and small-scale adaptations between divergent outcome teams. For RQ1, we compared SSMR strategy applications between MSOT and LSOT through three themes. Themes 1 and 2 focused on quantitative differences in SSMR strategies related to planning and task performance, comparing average statement occurrences (illustrated in Figure 2) and supporting findings with verbal interaction excerpts. Theme 3 examined qualitative differences in monitoring approaches for task progress and deliverables, supported entirely by verbal interaction evidence. For RQ2, we compared small-scale adaptations in SSMR strategies between MSOT and LSOT through two emergent themes (Themes 4 and 5). Both themes centered on

quantitative differences in adaptation frequency related to task planning and performance, as well as adaptation quality marked by occurrences of maladaptive traces (illustrated in Figure 3). Comparisons were strengthened through relevant verbal interaction excerpts demonstrating adaptive patterns.

Results

Descriptive statistics

The descriptive statistics of five teams show that teams applied various SSMR strategies, resulting in 85 SSMR strategy episodes containing 763 statements verbalized by team members during CPS (Table 2). The statistics reveal that SSMR strategies such as monitoring/controlling task performance and planning task performance were widely practiced, whereas monitoring/controlling task understanding and monitoring/controlling content understanding were less frequently observed. A similar trend emerged for adaptation traces: those targeting planning and executing task performance were widely exhibited, while adaptation traces targeting task and content understanding were less frequently exhibited.

Table 2

Descriptive Statistics of SSMR Strategy Episodes, Strategy Wise Statements, and Statements Targeting Different Regulation Phases (Aligned With Task)

	Team 1	Team 3	Team 4	Team 2	Team 5	Total
Total statements during CPS	1819	2119	2101	3179	1632	10850
Extracted SSMR strategy episodes	12	22	18	22	11	85
Total statements in SSMR strategy Episodes	149	183	137	193	101	763
Statements under different type of SSMR strategy						
Monitoring/control task understanding	13	25	11	20	5	74
Monitoring/control content understanding	16	3	21	5	19	64
Planning task performance	36	65	47	61	43	252
Monitoring/control task performance	84	90	58	107	34	373
Statements in adaptation traces during SSMR strategy						
Adaptation statements targeting task understanding	16	0	6	9	0	31
Adaptation statements targeting content understanding	0	3	13	3	9	28
Adaptation statements targeting planning task performance	26	33	13	8	0	80
Adaptation statements targeting task performance	32	10	25	23	9	99

Note: Team 1,3, and 4 shown in green color are more successful outcome teams (MSOT), and Team 2 and 5 shown in pink color are less successful outcome teams (LSOT).

For RQ1

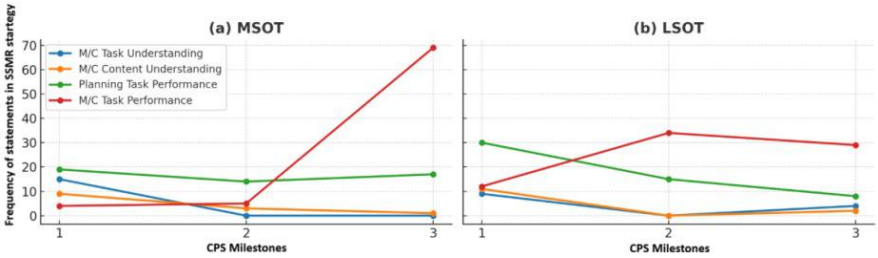
The comparative analysis of the SSMR strategy application by MSOT and LSOT are presented using themes below. The phase-wise differences in strategies are shown in the figure 2.

Theme 1: Consistency in planning strategy

The Figure 2 shows milestone-wise planning strategy statements exhibited by MSOT (19, 14, 17) and LSOT (30, 15, 8) respectively, indicating that MSOT applied planning strategies more consistently across all CPS phases compared to LSOT. Considering the longer duration and diverse subtasks of milestone 3, teams required consistent application of effective planning strategies. The declining planning applications by LSOT aligned with their task performance. Verbal interactions from LSOT (team 2) in milestone 3 revealed team members discussing remaining tasks while M2 read from the task sheet. M4 asked *"Hey M2, can you tell me what that (last milestone) is?"* and M2 responded *"its mapping one to one.. (solution to problem statement)."* This indicates lack of shared understanding about task goals even in the last milestone. Further, though M1 and M3 both worked on a shared coding module, M1 assigned a task to M3 by saying *"I need this type of program (shows snippet) in C++ format (for testing),"* showing an approach of dividing tasks into smaller units and assigning them to others instead of voluntarily taking responsibility. Conversely, verbal interaction from MSOT (team 1) showed when team members discussed a difficulty while implementing their solution's output module in milestone 3, M3 suggested to M2 *"Why will you send response to editor screen, otherwise try sending it in error (translation) screen.."* and M2 took up the responsibility by saying *"Ok. I will try that."* This indicates necessary shared-ness of internal task elements and willingness to take up responsibility voluntarily.

 © ISLS

Figure 2
The Milestones-Wise Application of SSMR Strategy (Average) are Compared Between MSOT and LSOT



While applying planning strategies, MSOT members took up task responsibilities voluntarily, attributed to planning team coordination, whereas LSOT members engaged more in assigning tasks to others and establishing shared understanding even in the final CPS phase.

Theme 2: Prioritization of task performance strategies in milestone 3

Milestone-wise task performance strategy statements exhibited by MSOT (4, 5, 69) and LSOT (12, 34, 29) differ substantially (Figure 2). MSOT deliberately prioritized task performance strategies in milestone 3 where actual solution implementation was initiated, whereas LSOT emphasized task performance strategies in milestones 1 and 2 but could not execute them on a large scale in milestone 3 where they were most needed due to longer duration and diverse subtasks. The prominent strategies applied by MSOT were questioning another person's idea/reasoning, asking/giving advice to peer, gauging task strategy with peers, evaluating task strategy, and monitoring current task performance. In contrast, LSOT primarily applied gauging task strategy with peers, evaluating task strategy, and monitoring current task performance. This indicates non-deliberative application of task performance strategies by LSOT in milestone 3, emphasizing gauging, monitoring and evaluating, whereas MSOT demonstrated deliberative prioritization of task performance strategies in milestone 3 with thoughtful implementation through refining and advising along with gauging, monitoring and evaluating.

Theme 3: Monitoring strategies for task progress and deliverables

A meaningful difference was observed in teams' monitoring behavior targeted towards task progress and deliverables. MSOT monitored and reviewed overall progress and milestone-wise deliverables more frequently and actively than LSOT. Teams perceived milestone-wise deliverables as goals and prepared various SSMR strategies to achieve them. While monitoring progress and deliverables, MSOT frequently applied strategies such as placing performance on shared or visible location, identifying points of discussion/decision, reviewing progress of task, monitoring current task performance, and evaluating task strategy. Conversely, LSOT applied monitoring current task performance, evaluating task strategy, and reviewing progress (to get everyone on same page).Verbal interactions from MSOT (team 3) showed while everyone reviewed the final solution's nature using a system diagram on the board, M2 initiated discussion by saying *"So what we are going to do exactly. See (points to diagram) there is front-end in front of you,"* indicating placement of performance on shared location and identification of decision points. Conversely, while discussing existing solutions, M4 from LSOT (team 2) asked M2 *"help me in this, I am not understanding this.. existing (design) solutions means what we are doing currently (designing 3 ideas). same right?"* and M2 replied *"Yeah."* M4 and M2 understood the task of gathering existing solutions incorrectly, indicating LSOT's inaccurate task understanding, with implications seen in deliverables. This fragmented monitoring approach resulted in LSOT's shortfall in delivering outcomes such as reviewing existing solutions for the chosen problem and mapping solution prototype with problem statement, indicating thorough application of SSMR strategies by MSOT versus fragmented application by LSOT for monitoring task progress and deliverables.

For RQ2

The comparative analysis of the adaptation traces in SSMR strategies by MSOT and LSOT are presented using themes below. The milestones-wise differences in adaptations in strategies are shown in the figure 3.

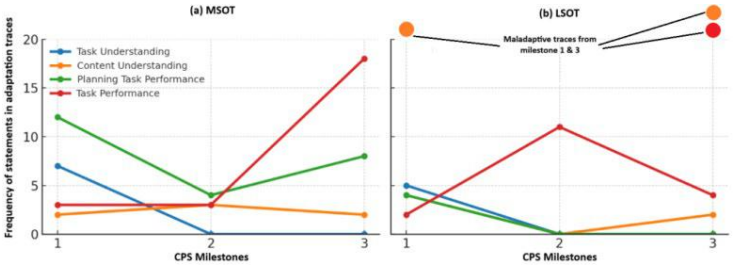
Theme 4: Magnitude of task planning and performance-related adaptations

There is substantial difference in average adaptation trace statements between MSOT and LSOT (Figure 3), particularly for adaptations targeted at planning and task performance strategies. For task planning adaptations, MSOT contains 12, 4, and 8 statements respectively across three milestones, whereas LSOT contains 4, 0, and

0. Notably, LSOT exhibited no adaptive changes in planning during milestones 2 and 3, which collectively account for substantial CPS duration.

Figure 3

The Milestones-Wise Adaptations (Average) in SSMR Strategy are Compared Between MSOT and LSOT



Note: The maladaptive traces from LSOT are shown using circle (containing same color as trend line color)

For task performance adaptations, MSOT contains 3, 3, and 18 statements respectively across three milestones, whereas LSOT contains 1, 11, and 4. This indicates MSOT's evident adaptive regulation behavior for task planning and task performance strategies compared to LSOT.

Theme 5: Regulatory adaptation quality (occurrences of maladaptive regulation)

Small-scale adaptations can be desired (adaptive) or undesired (maladaptive). LSOT (teams 2 and 5) exhibited three traces of maladaptive regulation, collectively accounting for 10 statements. These maladaptive statements targeted SSMR strategies for content understanding and task performance in separately spaced SSMR strategy episodes across milestones (Figure 3a & b). Table 3a shows one episode from LSOT (team 2) applying SSMR strategy related to content understanding, containing a trace of maladaptive small-scale adaptation. Conversely, MSOT displayed no maladaptive regulation behavior throughout SSMR strategy application.

Table 3a and b juxtaposes two episodes from LSOT and MSOT containing maladaptive and adaptive traces respectively. LSOT (team 2) members discussed writing a problem statement using the given problem scenario. Team 2 exhibited incorrect understanding of the task and deliverable (statements 11-12) and maladapted upcoming actions based on that (statements 13-15). Whereas, MSOT (team 1), while discussing an important internal component of the collaborative coding task, exhibited correct understanding of the task and deliverable (statements 107-108) and adapted upcoming actions accordingly (statements 109-111). This indicates differences in adaptation quality and its impact on task performance between LSOT and MSOT.

Table 3 a

The Maladaptive Trace from LSOT and Adaptive Trace from MSOT is Highlighted from Respective SSMR Episode

LSOT (Team 2)	Metacognitive skill	SSMR strategy	Adaptation
Context: In milestone 1, members of team 2 were discussing how to write a problem statement from given problem scenario			
(Statement 11) M1- I don't think so it's a (different than problem statement). It is the problem statement I feel.	(M-CM)	M/C content understanding Social questioning	-
(Statement 12) M3- Yeah, that is PS actually	(O-TA)	Response	-
(Statement 13) M4- No, They have given the scenario. So paraphrasing it only, we can make a problem statement.	(M-CM) (O-TA)	(Gauge content understanding with peers)	Mal-adaptive trace of small scale adaptation (statement 13-15)
(Statement 14) M1- Ok, like that, So this is situation, where we can..	(M-CM)	(Gauge content understanding with peers)	
(Statement 15) M4- yeah. Just remove scenario.	Other	Response	

Note- Metacognitive skills: (M-CM: Monitoring-comprehension monitoring), (O-TA: Orienting-task analysis), (Other-cognitive skill), (M-MC: monitoring collaboration), (P-SP: strategic planning).

Table 3 b

The Maladaptive Trace from LSOT and Adaptive Trace from MSOT is Highlighted from Respective SSMR Episode

 © ISLS

MSOT (Team 1)	Metacognitive skill	SSMR strategy	Adaptation
Context: In milestone 3, when team members of team 1 were discussing important internal component of the collaborative coding task			
(Statement 107) M3- Tell me how is your API? Show me that. How you need input?	(M-CM)	M/C task performance Gauge task strategy with peers	-
(Statement 108) M2- its there. In compile..its there completely (explaining.)	Other	Response	-
(Statement 109) M3- so tell me one thing. You are going to integrate or I should do?	(M-CM)	Gauge task strategy with peers	Adaptive trace of small scale adaptation (statement 109-111)
(Statement 110) M2- Shall I integrate?	(M-CM)	Gauge task strategy with peers	
(Statement 111) M3- Ohk, So my work is dependent on your work. So you do it first.	(M-MC) (P-SP)	Identify parts of the task which require different strategies	

Discussion and conclusion

We investigated SSMR strategies and small-scale adaptations enacted by divergent outcome teams (MSOT and LSOT) during ill-structured CPS. Using a mixed-methods approach with comparative thematic analysis, we revealed differences between SSMR strategies and. Findings highlight that MSOT consistently applied planning strategies for team coordination, while LSOT demonstrated inconsistent application. In milestone 3, LSOT showed non-deliberative application of task performance strategies, whereas MSOT demonstrated deliberative prioritization. For monitoring task progress and deliverables, MSOT exhibited thorough application of SSMR strategies compared to LSOT's fragmented approach. Additionally, MSOT demonstrated adaptive regulation behavior for task planning and performance strategies compared to LSOT which have impact of outcome.

Analysis for the first research question revealed vital differences between MSOT and LSOT. Theme 1 highlighted differences in consistency of task planning strategies quantitatively (Figure 2) and qualitatively. The constructive impact of micro-strategies like voluntarily taking up responsibility to maintain team coordination on MSOT's task performance is evident, whereas the adverse impact of inconsistent planning strategies like arbitrarily assigning small task responsibilities was visible on LSOT's task performance. LSOT lacked shared-ness. Theme 2 showed differences in number and enactment of task performance strategies. In milestone 3, MSOT handled these challenges thoughtfully using refining and advising strategies along with gauging, monitoring, and evaluating strategies, while LSOT applied gauging, monitoring, and evaluating strategies in a limited way. Theme 3 showed differences in monitoring task progress and deliverables. Further, MSOT placed performance on shared or visible location, identified points of discussion/decision, and reviewed task progress. LSOT applied monitoring and evaluating task performance strategies for subtasks instead of overall CPS progress and deliverables. Additionally, LSOT reviewed progress to maintain shared-ness page rather than reviewing overall CPS progress. The findings for RQ1 align with Lobczowski et al. (2021).

Analysis for second research question using theme 4 revealed differences in magnitude of task planning and performance-related adaptations. The ill-structured problem scenario required teams to adaptively plan and execute collaborative strategies considering dynamically changing task requirements. MSOT showed desired adaptive behavior, but LSOT failed to adapt for ill-structured CPS regarding SSMR strategy enactment. Theme 5 uncovered meaningful differences in small-scale adaptation quality. Maladaptive traces exhibited by LSOT had undesired impact on task performance, whereas MSOT exhibited no maladaptive behavior. Additionally, LSOT lacked uptake: they applied SSMR strategies and made adaptations, but weak uptake was visible post-adaptations across CPS milestones. Conversely, MSOT's uptake was strong as team members match the dynamic nature of subtasks and followed up on pre-decided courses of action. The findings for RQ2 align with Sobocinski et al. (2021), Sobocinski et al. (2019), and Järvelä et al. (2019). This study offers empirical evidence on SSMR strategy application and small-scale adaptations in divergent outcome teams during long-duration CPS tasks, advancing SSMR literature. These insights into SSMR strategies and adaptations highlight important implications for scaffolding comprehensive strategy use and minimizing maladaptive SSMR. However, limitations include the small, closed-lab sample size, potentially limiting generalizability, and a narrow focus that excluded the influence of individual mediating variables and the interplay of multiple regulatory forms. Future research should address these by using larger-scale, multimodal designs.

Endnotes

(1) The metacognitive skill coded in the first phase provides more fine grained details about SSMR strategy.
(2) The SSMR strategy code contains the target of the strategy which is same for target of the adaptations.

References

Ader, M., Hassane, S., van Bruggen, J., & Vermeulen, M. (2023). Comparing metacognitive regulation and socially shared metacognitive regulation in face-to-face and online learning settings in ill-structured problem-solving. Learning, Culture and Social Interaction, 39, 100684.

Çoklar, A. N., & Akçay, A. (2018). Evaluating Programming Self-Efficacy in the Context of Inquiry Skills and Problem-Solving Skills: A Perspective from Teacher Education. World Journal on Educational Technology: Current Issues, 10(3), 153-164.

Hadwin, A., Järvelä, S., & Miller, M. (2018). Self-regulation co-regulation and shared regulation in collaborative learning environments. In D. H. Schunk & J. A. Greene (Eds.), Handbook of self-regulation of learning and performance. (2nd ed., pp. 83–106). Routledge.

Iiskala, T., Vauras, M., Lehtinen, E., & Salonen, P. (2011). Socially shared metacognition of dyads of pupils in collaborative mathematical problem-solving processes. Learning and Instruction, 21(3).

Iiskala, T., Volet, S., Lehtinen, E., & Vauras, M. (2015). Socially shared metacognitive regulation in asynchronous CSCL in science: Functions, evolution and participation. Frontline Learning Research, 3(1), 78-111. https://doi.org/10.14786/flr.v3i1.159

Isohatala, J., Naykki, P., & Jarvela, S. (2019). Cognitive and socio-emotional interaction in collaborative learning: Exploring fluctuations in students' participation. Scandinavian Journal of Educational Research,

Järvelä, S., Järvenoja, H., & Malmberg, J. (2019). Capturing the dynamic and cyclical nature of regulation: Methodological Progress in understanding socially shared regulation in learning. International journal of computer-supported collaborative learning

Järvelä, S., Järvenoja, H., Malmberg, J., & Hadwin, A. F. (2013). Exploring socially shared regulation in the context of collaboration. Journal of Cognitive Education and Psychology, 12(3).

Järvelä, S., Malmberg, J., Sobocinski, M., & Kirschner, P. A. (2021). Metacognition in collaborative learning. In the International handbook of computer-supported collaborative learning (pp. 281-294). Springer

Järvenoja, H., Järvelä, S., & Malmberg, J. (2020). Supporting groups' emotion and motivation regulation during collaborative learning. Learning and Instruction, 70, 101090.

Joffe, H. (2012). Thematic analysis. In D. Harper & A. R. Thompson (Eds.), Qualitative research methods in mental health and psychotherapy

Lobczowski, N. G., Lyons, K., Greene, J. A., & McLaughlin, J. E. (2021). Socially shared metacognition in a project-based learning environment: A comparative case study. Learning, Culture and Social Interaction, 30, 100543

Malmberg, J., Järvelä, S., Järvenoja, H., & Panadero, E. (2015). Promoting socially shared regulation of learning in CSCL: Progress of socially shared regulation among high- and low-performing groups. Computers in Human Behavior, 52, 562–572.

Mayring, P. (2015). Qualitative content analysis: Theoretical background and procedures. Approaches to qualitative research in mathematics education: Examples of methodology and methods, 365-380

Meijer, J., Veenman, M. V. J., & van Hout-Wolters, B. H. A. M. (2006). Metacognitive activities in text-studying and problem-solving: Development of a taxonomy. Educational Research and Evaluation.

Seiradakis, E. & Spantidakis, I. (2019). Socially-Shared Metacognitive Regulation Episodes between Undergraduate Engineering Students during a Collaborative Genre Analysis Task in an English for Academic Purposes Course. English Language Teaching.

Schoor, C., Narciss, S., & Körndle, H. (2015). Regulation during cooperative and collaborative learning: A theory-based review of terms and concepts. Educational Psychologist, 50(2), 97–119. https://doi.org/10.1080/00461520.2015.1038540

Shin, Y., Lee, H., & Kim, W. (2025). Exploring Creative Problem-Solving in Computer-Supported Collaborative Learning: Focusing on Group Cohesiveness and Socially Shared Metacognitive Regulation. Thinking Skills and Creativity, 101771. https://doi.org/10.1016/j.tsc.2025.101771

Sobocinski, M., Järvelä, S., Malmberg, J., Dindar, M., Isosalo, A., & Noponen, K. (2020). How does monitoring set the stage for adaptive regulation or maladaptive behavior in collaborative learning?. Metacognition and Learning, 15(2), 99-127.

Sobocinski, M., Malmberg, J., & Järvelä, S. (2022). Exploring adaptation in socially-shared regulation of learning using video and heart rate data. Technology, Knowledge and Learning, 27(2), 385-404.

Tucker, T., Shehab, S., & Mercier, E. (2020). The impact of scaffolding prompts on the collaborative problem solving of ill-structured tasks by undergraduate engineering student groups. In 2020 ASEE Virtual Annual Conference Content Access

Ucan, S., & Webb, M. (2015). Social regulation of learning during collaborative inquiry learning in science: How does it emerge and what are its functions? International Journal of Science Education, 37(15), 2503–2532.

Vauras, M., Volet, S., & Iiskala, T. (2021). Socially-shared metacognitive regulation in collaborative science learning. Trends and prospects in metacognition research across the life span: A tribute to Anastasia Efklides.

Vogel, F., Weinberger, A., Hong, D., Wang, T., Glazewski, K., Hmelo-Silver, C. E., ... & Noroozi, O. (2023). Transactivity and Knowledge Co-Construction in Collaborative Problem Solving. In Proceedings of the 16th International Conference on Computer-Supported Collaborative Learning-CSCL 2023, pp. 337-346. International Society of the Learning Sciences.

Winne, P. H., & Hadwin, A. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), Metacognition in educational theory and practice. (pp. 277–304). Routledge

Acknowledgments

We sincerely thank all the participants for their valuable time and insights. Their contributions were essential to the success of this study.

International Society of the Learning Sciences

Close Encounters: Teacher Proximity and Student-Reported Emotions While Using ST Math

Sotheara Veng, Yiqin Chao, Rosa Mykyta-Chomsky, Teomara Rutherford
sotheara@udel.edu, caoemma@udel.edu, rmykytac@udel.edu, teomara@udel.edu
University of Delaware

Abstract: Students' academic emotions are critical to learning (Pekrun et al., 2002; Tan et al., 2021) and are linked to teacher actions (Meyer & Turner, 2007), some of which require close proximity. Prior studies emphasize the importance of proximity in shaping classroom environments, student behavior, and math anxiety (Shores et al., 1993; Quinnell, 2022). Yet, research on proximity is limited, especially in technology-mediated classrooms where teachers may not necessarily be in close proximity to students. This study examined the moment-to-moment, bidirectional relationship between student-reported emotions and teacher proximity during ST Math sessions. Using multilevel vector autoregression on experience sampling data from 92 elementary students, we found that frustration was associated with closer subsequent proximity, but it is the opposite for boredom. Conversely, closer proximity predicted lower subsequent boredom. These findings suggest that teacher proximity may play an important role in managing students' emotions in technology-mediated settings.

Introduction

Students' academic emotions play an important role in learning, as they are linked to motivation, self-efficacy, achievements, and self-regulation (see Schutz & Lanehart, 2002; Tan et al., 2021). These emotions are associated with teachers' actions in the classroom (Lei et al, 2018). Moreover, teacher-student proximity, the physical distance between teachers and students in the classroom, is also linked to student participation, engagement, and performance (Dong et al., 2021; Marshall & Losonczy-Marshall, 2010). Teachers' proximity to students can shape how the students perceive their teacher's presence. For example, being nearby may convey care and togetherness (Plante & Asselin, 2014). Moreover, when teachers are physically closer to students, students tend to participate more actively in discussions and feel greater engagement with course content (Meeks et al., 2013). Given these links, it is important to understand what drives proximity and how proximity may lead to outcomes. The relationship between a teacher's physical closeness and student engagement or performance may be complex. There may be variation between teachers and students in how teacher proximity is associated with emotions; for example, a teacher standing nearby may elicit anxiety in some students, which can be negatively associated with their participation or achievement (Brumariu et al., 2023; Pekrun et al., 2017). Examining students' emotional responses may provide a more complete understanding of how physical distance in the classroom is linked to learning. Furthermore, the association may also operate in reverse. For instance, certain students' emotions, such as signs of frustration or disengagement, may trigger teachers to move closer to those students to provide support (Guther et al., 1995). Investigating bidirectional links between proximity and emotions will shed light on the dynamics of this important teacher-student classroom interaction.

Within the current study, we investigate these associations in the context of a learning technology. Student use of learning technologies in the classroom presents a unique wrinkle to understanding student/teacher interactions, as students often tend to work independently on technology with minimal teacher guidance (e.g., Weng et al., 2020). However, teachers continue to play an important role in supporting and facilitating student engagement with these technologies by providing assistance when needed (Negoescu & Boştină-Bratu, 2016). Teachers can support students' emotional experiences with technology by checking in, encouraging them, and providing scaffolding to students with negative emotions, such as confusion and boredom. Understanding how teachers' movements within classrooms relate to students' emotions while they interact with technology can provide insight into the teacher-technology-student ecology and allow for the development of practical recommendations for teachers in supporting their students' emotions, engagement, and performance. In this paper, we aimed to answer the following research question:

1. How is teacher-student physical proximity associated with subsequent students' reported emotions while using ST Math?
2. How are student-reported emotions associated with subsequent teacher-student physical proximity during ST Math sessions?

 © ISLS

Theoretical framework

Within the current study, we explore these associations through the lens of the cognitive-motivational model of academic emotions (Pekrun et al., 2000), sociocultural theory (Vygotsky, 1978), and proxemic theory (Hall, 1968). These theories provide a comprehensive framework for understanding how to examine the complex interplay between emotions, social interactions, and physical proximity in the classroom environment. These perspectives together allow us to conceptualize student emotions as dynamically shaped by both internal appraisals and external classroom interactions, including teachers' physical proximity to students.

Academic emotions

Academic emotions, as conceptualized by Pekrun et al. (2002), refer to the specific emotions students feel during academic activities. These emotions can include both positive feelings (e.g., enjoyment, interest) and negative feelings (e.g., anxiety, boredom, and frustration), all of which can impact students' motivation, engagement, and learning outcomes (Pekrun et al., 2002). Positive emotions can improve motivation, promote flexible learning strategies and self-regulation, allowing more cognitive resources for engaging with a task (Pekrun et al., 2002). On the other hand, negative emotions are believed to decrease motivation and cause students to rely on more restrictive learning approaches, like basic rehearsal and algorithmic procedures (Pekrun et al., 2002). Moreover, negative emotions can reduce the cognitive resources and self-regulation available, because when students experience these negative emotions, they may have less mental energy to focus on learning tasks, less motivation to engage with the material, and may struggle with effectively regulating their own learning process (Pekrun et al., 2002).

Moreover, emotions can also be distinguished by their level of physiological activation: positive activating (e.g., enjoyment, excitement), positive deactivating (e.g., relief, contentment), negative activating (e.g., frustration, anxiety), and negative deactivating (e.g., boredom, hopelessness; Pekrun, 2006; Pekrun, 2024). Activating emotions are characterized by heightened arousal that mobilizes cognitive and behavioral resources toward the task at hand (Pekrun, 2006).

Deactivating emotions, by contrast, are characterized by low arousal that tends to reduce engagement and effort (Pekrun, 2006). Therefore, activating emotions may be more visible and therefore more detectable compared to deactivating emotions, such as boredom, which are more passive. Positive activating emotions are associated with increased motivation, flexible and deep learning strategies, and sustained cognitive engagement (Pekrun et al., 2002). Positive deactivating emotions (e.g., relief, contentment) are linked to lower urgency to engage further, and thus have weaker or more mixed associations with subsequent learning effort (Pekrun, 2024). Negative activating emotions (e.g., frustration, anxiety) are characterized by heightened arousal and continued effortful engagement with the source of difficulty; although they can motivate persistence in some circumstances, they typically reduce available cognitive resources and promote more rigid, surface-level learning approaches (Pekrun et al., 2002). Negative deactivating emotions (e.g., boredom, hopelessness), on the other hand, are characterized by low arousal, passivity, and withdrawal from the task altogether, and consistently show the most detrimental effects on motivation, self-regulation, and achievement (Pekrun, 2024).

Sociocultural theory

Although academic emotions are often conceptualized at the individual level, sociocultural theory provides a lens for understanding how these emotions are shaped through social interactions in the classroom. Sociocultural theory, as developed by Vygotsky (1978), emphasizes the importance of social interactions and cultural contexts in shaping individuals' cognitive development and learning. The theory suggests that learning is inherently a social and cultural process. From a sociocultural perspective, emotions can be viewed as both individual and social phenomena, without separating them into different realms or aspects (Sánchez, 2014). In other words, the sociocultural perspective implies examining emotions as socially situated rather than purely internal states. This viewpoint suggests that the interactions between teachers and students in the classroom play an important role in shaping students' emotional experiences. An ecological dynamic systems synthesis of classroom emotion research by Savina and Fulton (2024) reinforces this position by identifying three immediate contexts in which emotions are generated and shaped in the classroom. The first is the emotionally expressive environment, in which teachers and students continuously elicit and respond to each other's emotional expressions, including through nonverbal cues, such as physical proximity, facial expressions, and body orientation, which create conditions for emotional contagion and affective mimicry. The second is the learning and instructional context, in which the nature of tasks, students' learning behaviors, and teachers' instructional actions intersect to produce emotions; for instance, unclear instruction or tasks that exceed students' perceived control can give rise to frustration and boredom, whereas supportive instructional behaviors promote positive emotions. The third is the relational context, in which the

quality of teacher-student relationships shapes students' emotional baseline. Students in warm, close relationships with their teachers experience more positive emotions and less anxiety and boredom, but conflictual or distant relationships are associated with more negative emotional experiences. These three contexts can be active simultaneously during a classroom session where students' emotional experiences are shaped by the expressiveness of the classroom environment, the instructional demands of the tasks, and the quality of their moment-to-moment interactions with their teacher, including how physically close that teacher is to them.

Proxemic theory

Within these social interactions, teacher-student physical proximity serves as an important component of classroom communication. Proxemic theory, as proposed by Hall (1968), examines the spatial relationships and distances between individuals in social interactions. According to the theory, the physical distance between individuals influences their interpersonal dynamics (Hall, 1968; Miller, 1988). For example, in a classroom setting, proximity between a teacher and students is linked to the quality of interactions, perceptions of support, and students' emotional experiences (Miller, 1988). Proximity is likely to have a role in students' feelings in class (Hall, 1968; Miller, 1988), attitudes toward academic tasks (Dong et al., 2021; Marshall & Losonczy-Marshall, 2010), and classroom participation (Meeks et al., 2013). Shapiro et al. (2024) proposed that teachers' movement through the classroom space reflects and enacts their responsiveness to students' needs, ideas, and emotional states. Moreover, research on teacher nonverbal immediacy, a construct that operationalizes physical and psychological closeness between teachers and students, has found that physical proximity and body orientation are associated with students' situational emotional engagement (De Loof et al., 2025) and that teacher immediacy behaviors more broadly predict student academic outcomes (Özdaş, 2022).

The framework of different zones of proximity was developed by Hall (1968), which examines how people perceive and use physical and social space across different cultures, including "intimate," "personal," "social," and "public" zones. Miller (1988) viewed proximity as a non-verbal cue that can communicate teacher attitudes and influence student performance and feelings. Kale (2008) adopted Hall's proximity zones as a coding scheme and found that classroom interactions were more likely to occur at certain proximity zones. Most other prior work examines physical proximity based on classroom seating arrangements instead (e.g., front row, back row) to measure its association with students' performance in college-level courses with little instructor movement (e.g., Marshall & Losonczy-Marshall, 2010; Meeks et al., 2013). In the context of K-12 classrooms, more teacher movements may be expected to occur. Therefore, we adopted Hall's (1968) framework of proximity zones. In our study, we examined the bidirectional relationship between physical proximity and student-reported emotions while students use the math game, ST Math.

Literature review

Academic emotions in the classroom and in technology environments

Prior research has indicated that students' academic emotions are linked to learning processes and outcomes, including motivation, engagement, and achievement (Pekrun et al., 2012; Tan et al., 2021). These emotions vary across individuals and contexts, shaped by factors such as students' prior experiences, self-concept, and perceptions of task difficulty and value (Goetz et al., 2010). Emotions are also domain-specific. For example, students may experience different emotional patterns in mathematics compared to language-based subjects due to differences in perceived difficulty, value, and self-efficacy (Goetz et al., 2011). Classroom environments may further shape emotional experiences. Instructional practices, peer dynamics, and classroom climate can influence whether students experience more activating emotions (e.g., excitement) or deactivating emotions (e.g., boredom; Goetz et al., 2011; Pekrun, 2006). In technology-mediated learning environments, emotional experiences may differ from those in traditional classrooms. Although some studies suggest that digital learning environments can promote positive emotions (Loderer et al., 2020), students may still experience negative emotions, such as boredom or confusion, particularly when tasks are misaligned with their skill levels (Acee et al., 2010). These findings highlight that technology does not eliminate the need for instructional and emotional support. Despite this growing body of research, less is known about how teachers' in-the-moment behaviors, such as physical proximity, relate to students' emotional experiences during technology use. Understanding these dynamics is critical for clarifying how teacher presence continues to shape learning in increasingly technology-mediated classrooms.

Teacher-student interactions

Teacher-student interactions may be an important part of students' school experiences, as they serve a variety of purposes in the classroom, including instructional, organizational, and emotional support (Pianta & Hamre, 2009). For example, teachers can offer instructional support by giving clear feedback to students, creating opportunities for conceptual thinking, and modeling new vocabulary. Studies have shown that students who receive more support from teachers have more positive academic emotions (e.g., enjoyment and interest) and fewer negative emotions (e.g., anxiety and boredom) (e.g., Ahmed et al., 2010). Moreover, instructional behaviors, the ways instructional support is carried out, were also associated with students' emotions. These behaviors can directly shape students' perceptions of control and value, which have been shown to be linked to emotional experiences (Bieg et al., 2013; Pekrun, 2006). For instance, providing clear explanations and becoming more engaging when teaching can help promote positive control and value appraisals, respectively (Bieg et al, 2013). Prior work confirmed that both of these instructional behaviors are related to students' positive emotions in different subjects, such as mathematics and physics (e.g., Goetz et al., 2013). Conversely, students are more likely to experience negative emotions when teachers provide unclear explanations or communication (Mazer, 2014). Teacher-student proximity can play a role in this. Being in closer proximity to students may allow teachers to implement more effective instructional practices as they can be more audible and become attuned to gaps in student understanding, which can promote students' positive emotions.

Alternatively, teachers' actions can also be influenced by students' emotions, as teachers respond to the emotional needs of students in their classrooms (Jennings & Greenberg, 2009). Teachers may provide emotional support to students by being attentive, understanding, caring, and mindful of students' needs and interests (Pianta & Hamre, 2009). Emotional support has been shown to be beneficial in math classes, with students showing higher emotional engagement, academic effort, and social engagement when they feel the emotional support and care from their teachers (Sakiz et al., 2012). The term "Emotional Scaffolding" has been coined to describe such support (Meyer & Turner, 2007). Stemming from the scaffolding component of Vygotsky's zone of proximal development (Vygotsky, 1978), emotional scaffolding is defined as teachers fostering students' positive emotional experiences to benefit learning and growth (Meyer & Turner, 2007). The Meyer and Turner (2007) emotional scaffolding framework proposed that students need to receive temporary support from teachers' scaffolding of their emotions when they are in urgent need of that assistance. Therefore, teachers may approach students exhibiting frustration, anxiety, or low confidence to provide emotional scaffolding.

The physical proximity between teachers and students may be one observable manifestation of this emotional support. During class, teachers navigate closeness to or distance from individual students. For example, teachers may intentionally approach students they believe are struggling or feeling frustrated to provide them with targeted assistance. Similarly, students exhibiting signs of disengagement may cause teachers to move closer to engage them (Etscheidt et al., 1984; Guther et al., 1995; Weaver et al., 2020). However, as the teachers move closer to some students, teachers will be at a farther physical distance from some other students. This may inadvertently limit emotional support for other students. Our work examined in-the-moment teacher-student interactions through the lens of physical proximity. This perspective highlights the need to observe and analyze the physical distance between teachers and students during real-time interactions. Mapping cross-time changes in proximity to student emotions can expand the area of optimal teacher responsiveness.

Methods

Positionality statements

The author team includes graduate students and a faculty member with backgrounds in education and learning sciences. Although two were educated in the United States (one in U.S. public schools), none were educated in the community in which the study took place; three of the authors have been K-12 teachers. The faculty co-author's research group has a long-running relationship with the study school district that includes a number of studies and immersion in classrooms (e.g., volunteering). All first three authors were directly involved in classroom observations, and two were involved in proximity coding. These experiences, both in and out of the specific study context, have shaped how we approached the research and likely influenced our interpretation of results.

Context

The current study is embedded in a larger project studying the mathematics learning software, ST Math, which is used by over one million students across the United States. The project is funded both by the authors' university and by the National Science Foundation. Project researchers collected classroom data as students interacted with

the mathematics learning platform, ST Math, as implemented during the 2022-2023 school year in a school district in the mid-Atlantic region of the United States. ST Math is an educational game-like online supplemental curriculum that students use on their computers or tablets throughout the school year, typically for an average of 90 minutes per week. ST Math use has been shown to be positively associated with mathematics scores and student self-beliefs (Rutherford et al., 2019).

Sample and participants

The participants included 92 students from grades four and five across two public elementary schools within the classes of 10 teachers, each teaching one class, and two additional co-teachers; the co-taught class was treated separately for analysis. Eight teachers taught at School A and four teachers at School B. The school district comprises 7,826 students enrolled in public schools, 52.1% of whom are boys. The district serves a relatively diverse student body; 92% of students identify as a single race, and 8% identify as multi-racial. Of the single-race responses, 33% identify as White, 43.5% as Black, 3.5% as Asian, and 11.7% as other races (National Center for Education Statistics, 2023).

The study adhered to ethical guidelines for research involving human subjects; an IRB was secured from the authors' university. All participants voluntarily participated in the study and provided informed consent. For students, parental or guardian consent was also obtained.

Participating students typically used ST Math either in group rotation or individually at their desks as a whole class. The group rotation was based on predetermined schedules given by the teacher, typically shifting every 15 minutes. When students were not on an ST Math rotation, they would work with the teacher or an assistant on other math-related content. Teachers would sometimes walk around the class giving students individual attention; some teachers asked that students with questions approach their desks for help, forming a line when other students were already being helped. In other cases, teacher assistants helped with student questions related to ST Math while the teachers either conducted other work or assisted other students. In both group rotation and whole class approaches, students used ST Math individually.

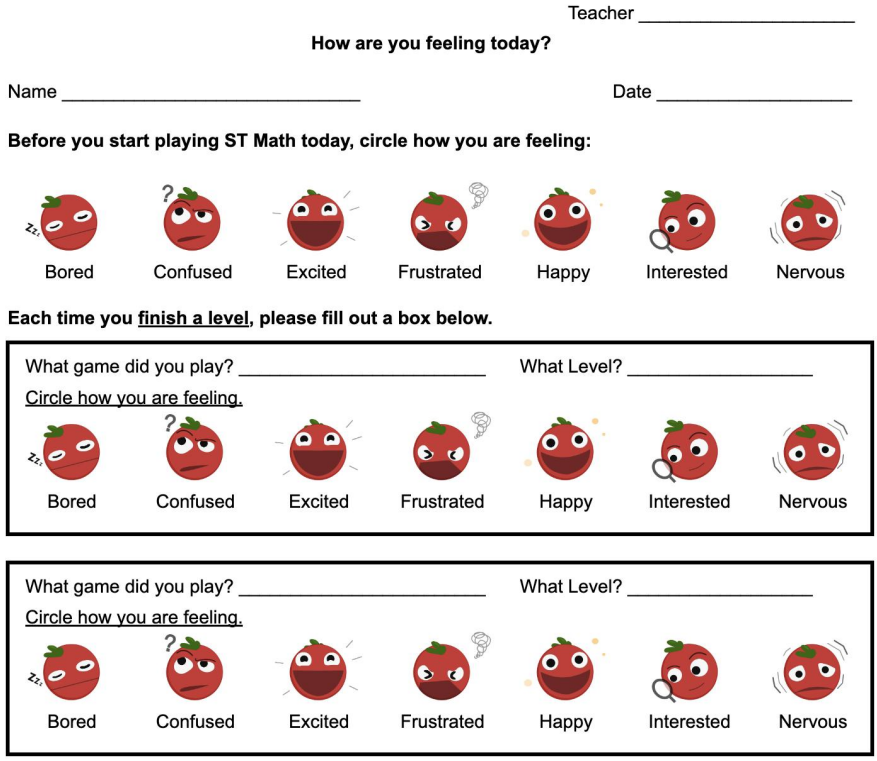
Data collection

Each observation, conducted by teams of three undergraduate and/or graduate research assistants, captured teacher proximity and actions and lasted 18–51 minutes. The observation involved creating classroom layouts (or using pre-created layouts), tracing teacher movement with timestamps, and synchronizing to map locations. These actions were recorded in real-time as they occurred by researchers on paper maps of the classroom. Senior researchers trained each team through on-site supervision for at least three sessions. Hands-on training continued until teams could reliably conduct observations independently.

Students were given experience-sampling method (ESM; Csikszentmihalyi & Larson, 2014) surveys that consisted of tomato-like emojis called "tomojis," which visually represented the emotion listed on a tomato's face. To capture emotional variation, experience sampling measures were administered at a fixed interval and at the beginning of each observation. Prior to data collection, students were trained on how to use the emotion form. Students were asked to fill out one set of emotion entries as soon as they received the survey, and up to eight additional sets based on how much time was observed in the lesson. The additional sets were timed every five minutes by researchers, at which time students were asked to choose as many emotions as they were feeling at the time (a maximum of seven) from the following: bored, confused, frustrated, nervous, excited, happy, interested in the form of Tomojis (see Rutherford et al., 2021, Figure 1). This allowed continuous tracking of academic emotions over time and in response to classroom events (see Hektner et al., 2007).

Figure 1
Tomoji Form

 © ISLS



Note. The figure shows the top part of the Tomoji ESM form, which was used for experience sampling of students' academic emotions. The bounded boxes represent entries. Students selected amongst tomato emojis representing seven emotions to indicate how they felt at the moment of each ESM survey prompt.

Data analysis

Teacher-student proximity data were coded using Hall's (1968) framework relevant to each student with consent who was observed (see Table 1). Two researchers independently coded the proximity data to reflect the appropriate category at each observation timestamp. Only teacher actions within two minutes before/after ESM surveys were matched to surveys in order to capture the proximity closest in time to emotions (see Figure 2). After coding two to three teachers, researchers conducted consensus meetings to review codes, resolve any discrepancies, and agree on final consensus codes. In addition to the final consensus codes, the original codes from each coder were retained to calculate interrater reliability (IRR) using Cohen's kappa, with a value of k = 0.804 indicating a substantial level of agreement between coders (McHugh, 2012). The dataset consisted of 2,338 momentary observations from 92 students. At the within-person level (Level 1), the outcome variables included teachers' physical proximity to students, treated as a continuous variable, and seven binary-coded emotional states: bored, confused, excited, frustrated, happy, interested, and nervous.

We used multilevel vector autoregression modeling (Bringmann et al., 2013) to examine temporal dynamics among student emotions and teacher proximity. Multilevel vector autoregression was selected because it allows modeling of within-person temporal dynamics and bidirectional relationships while accounting for the nested structure of repeated observations within students. The model examined how variables predict themselves and each other over time while accounting for the nested structure of repeated observations within students (Bringmann et al., 2013). The model included two levels. At the within-person level (Level 1), we specified three types of temporal relationships for each student's moment-to-moment fluctuations: (a) autoregressive effects, where each variable predicted itself one moment later (e.g., bored at time t predicting bored at time $t + 1$), capturing temporal stability or "carry-over"; (b) cross-lagged effects where one variable at time t predicted a different variable at time $t + 1$ (e.g., proximity at time t predicting frustrated at time $t + 1$), testing temporal precedence among emotions; and (c) contemporaneous residual correlations among all variables at the same time point, accounting for shared variance not explained by lagged predictors. We specified bidirectional cross-lagged paths between all emotions and teacher-student proximity to examine whether emotions drew teachers closer and whether proximity influenced subsequent emotions. We also included theory-driven cross-lagged paths based on

 © ISLS

prior research (D'Mello & Graesser, 2012), specifically testing: confusion predicting frustration, frustration predicting boredom and nervousness, and bidirectional confusion-boredom relationships. However, for this paper, we only reported the findings on the reciprocal relationship between proximity and reported emotions. At the between-person level (Level 2), we estimated random intercepts for all variables, which allows individuals to differ in their average levels of each emotion and typical teacher proximity. For binary emotion variables, random intercepts were specified as thresholds in the probit model, representing individual differences in the propensity to endorse each emotion. We assessed the model convergence using the potential scale reduction (PSR) statistic, which compares variability between chains to variability within chains. PSR values close to 1.0 indicate that chains have converged to the same distribution. Following standard recommendations, we considered PSR < 1.10 as indicating adequate convergence (Gelman et al., 2013).

Results

We present results organized by research question (RQ). All models converged successfully (highest PSR = 1.096 at iteration 10,000).

Research question 1: Do student emotions predict teacher proximity?

Frustration, excitement, boredom, and confusion were associated with subsequent teacher proximity (Table 1). The largest association in our entire model was from frustration predicting proximity (b = 0.44, *SE* = 0.02, $p < .001$, β = 0.44). When students were one standard deviation (SD) more frustrated than their personal average, teachers moved 0.44 SD closer, which may be a response indicating teachers noticed and approached struggling students. Excitement was also associated with teachers being closer (b = 0.13, $p < .001$, β = 0.13), as was confusion (b = 0.08, $p < .001$, β = 0.08) and nervousness (b = 0.05, $p = .010$, β = 0.05), though these associations were smaller. Interestingly, boredom is negatively associated with proximity (b = -0.11, $p < .001$, β = -0.11), which means that on average, teachers tended to be farther away from those who reported feeling bored.

Table 1

Association between Student Emotions and Subsequent Teacher Proximity

Outcome	Predictor	b	SE	95% CI	β	*p*
Proximity	Frustrated(*t*-1)	0.435	0.023	[0.390, 0.481]	0.44	<.001
Proximity	Excited(*t*-1)	0.127	0.025	[0.080, 0.175]	0.128	<.001
Proximity	Bored(*t*-1)	-0.110	0.02	[-0.151, -0.071]	-0.111	<.001
Proximity	Confused(*t*-1)	0.076	0.02	[0.036, 0.116]	0.076	<.001
Proximity	Nervous(*t*-1)	0.048	0.021	[0.008, 0.089]	0.049	0.01
Proximity	Happy(*t*-1)	-0.004	0.022	[-0.046, 0.038]	-0.004	0.419
Proximity	Interested(*t*-1)	-0.006	0.026	[-0.056, 0.045]	-0.006	0.413

Note. b = unstandardized coefficient; SE = posterior standard deviation; CI = 95% credible interval; β = standardized coefficient. Positive coefficients indicate closer teacher-student proximity; negative coefficients indicate farther teacher-student proximity.

Research question 2: Does teacher proximity influence student emotions?

Teacher proximity was associated with subsequent boredom (*b* = -0.20, *SE* = 0.04, $p < .001$, β = -0.12). When teachers were 1 *SD* closer than typical, students' boredom decreased by 0.12 *SD* at the next moment. This may potentially suggest that teacher presence could maintain engagement or re-engage students. However, proximity was not associated with next-time-point frustration (*b* = -0.04, *SE* = 0.04, $p = .145$) or other individual emotions, although frustration was associated with subsequent proximity in the previous RQ.

Table 2

Association between Teacher Proximity and Subsequent Student Emotions

Outcome	Predictor	b	SE	95% CI	β	*p*
Bored	Proximity(*t*-1)	-0.196	0.042	[-0.280, -0.114]	-0.118	<.001
Confused	Proximity(*t*-1)	-0.079	0.045	[-0.168, 0.009]	-0.043	0.038

 © ISLS

Frustrated	Proximity(*t*-1)	-0.042	0.041	[-0.123, 0.037]	-0.021	0.145
Excited	Proximity(*t*-1)	-0.066	0.043	[-0.148, 0.019]	-0.024	0.062
Happy	Proximity(*t*-1)	-0.047	0.046	[-0.137, 0.044]	-0.022	0.157
Interested	Proximity(*t*-1)	0.046	0.044	[-0.040, 0.133]	0.023	0.153
Nervous	Proximity(*t*-1)	-0.038	0.04	[-0.117, 0.041]	-0.026	0.173

Note. b = unstandardized probit coefficient; SE = posterior standard deviation; CI = 95% credible interval; β = standardized coefficient. Only boredom showed a clear response to teacher proximity. The marginal effect on confusion (p = .038) should be interpreted cautiously, given the number of comparisons.

Discussion

Our study investigated the dynamic, bidirectional relationship between teacher proximity and student emotions during technology-based math lessons. The findings reveal an interplay where teachers were responsive to certain emotional cues (i.e., frustration, excitement, and confusion). Our results align with research on teachers' emotional support (Pianta & Hamre, 2009), which emphasizes the ability of effective educators to respond to student needs. We found that teachers moved closer to frustrated students, potentially demonstrating "emotional scaffolding," which may be important for supporting students in learning (Meyer & Turner, 2007). However, this proximity response did not extend to boredom. Teachers seemed to be farther from bored students, which may be concerning, as boredom has negative relationships with outcomes, such as academic performance and self-regulated learning. (e.g., Tze et al. 2013). This pattern may also align with the Control-Value Theory of achievement emotions (Pekrun, 2006), which distinguishes between activating and deactivating emotions. Activating emotions (e.g., excitement) often involves visible effort and engagement, which perhaps makes it easier for teachers to detect and interpret them as cues for instructional or emotional support (D'Mello & Graesser, 2012). In contrast, boredom is a deactivating emotion typically characterized by passivity and disengagement, which may not elicit immediate teacher response.

Interestingly, we found that closer teacher presence was associated with less subsequent student boredom. This may align with studies on the link between physical proximity and student engagement and participation and can non-verbally communicate a sense of care and togetherness (Plante & Asselin, 2014). These results may have direct implications for teachers' roles in technology-rich classrooms. Although technology facilitates individualized learning, the teacher's role as a facilitator providing support may be critical.

Even though this work highlights the importance of proximity, its limitations include a reliance on self-reported emotions, which may manifest differently than observed emotions. Additionally, proximity was operationalized as physical distance and does not capture the underlying meaning of teacher proximity (e.g., instructional support, or behavioral control). It is likely that the effect of proximity depends on teachers' actions and instructional approaches. More support from teachers, for example, can promote positive emotions and lower the likelihood of students' negative emotions (Ahmed et al., 2010). In our study, we observed teachers' actions as well, but there were not sufficient data points to be meaningfully analyzed and fit into our models within the current study, which should be examined in future research.

Conclusion

This study contributes to understanding the association between teacher proximity and student emotions during technology-mediated mathematics learning. Using moment-to-moment data and multilevel vector autoregression, we found that teachers moved closer when students reported being frustrated, confused, or nervous, but teachers tended to be farther away when students were bored, even though closer proximity predicted lower subsequent boredom. These results highlight the importance of teacher movement as both a behavioral and emotional scaffold that supports engagement in individualized digital learning contexts. Future research should further explore how teachers can strategically balance physical closeness across students and how other interactions can sustain emotional engagement in technology-rich classrooms.

References

Acee, T. W., Kim, H., Kim, H. J., Kim, J., Chu, H.-N. R., Kim, M., Cho, Y., Wicker, F. W., & The Boredom Research Group. (2010). Academic boredom in under- and over-challenging situations. *Contemporary Educational Psychology*, *35*(1), 17–27.

Ahmed, W., Minnaert, A., van der Werf, G., & Kuyper, H. (2010). Perceived social support and early adolescents'

achievement: The mediational roles of motivational beliefs and emotions. *Journal of youth and adolescence*, *39*, 36-46.
Bieg, M., Goetz, T., & Hubbard, K. (2013). Can I master it and does it matter? An intraindividual analysis on control–value antecedents of trait and state academic emotions. *Learning and Individual Differences*, *28*, 102–108.
Bringmann, L. F., Vissers, N., Wichers, M., Geschwind, N., Kuppens, P., Peeters, F., Borsboom, D., & Tuerlinckx, F. (2013). *A network approach to psychopathology: New insights into clinical longitudinal data. PLoS ONE*, *8*(4), e60188.
Brumariu, L. E., Waslin, S. M., Gastelle, M., Kochendorfer, L. B., & Kerns, K. A. (2023). Anxiety, academic achievement, and academic self-concept: Meta-analytic syntheses of their relations across developmental periods. *Development and psychopathology*, *35*(4), 1597-1613.
Csikszentmihalyi, M., & Larson, R. (2014). Validity and reliability of the experience-sampling method. *Flow and the foundations of positive psychology* (pp. 35-54). Springer, Dordrecht.
De Loof, H., Struyf, A., Boeve-de Pauw, J., & Van Petegem, P. (2025). Teacher nonverbal immediacy: A validation study of the TeNOI observation scale. *Scandinavian Journal of Educational Research*.
Dong, Z., Liu, H., & Zheng, X. (2021). The influence of teacher-student proximity, teacher feedback, and near-seated peer groups on classroom engagement: An agent-based modeling approach. *PLoS ONE*, *16*(1).
D'Mello, S., & Graesser, A. (2012). Dynamics of affective states during complex learning. *Learning and Instruction*, *22*(2), 145–157.
Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis* (3rd ed.). CRC Press.
Goetz, T., Cronjaeger, H., Frenzel, A. C., Lüdtke, O., & Hall, N. C. (2010). Academic self-concept and emotion relations: Domain specificity and age effects. *Contemporary Educational Psychology*, *35*(1), 44-58.
Goetz, T., Frenzel, A. C., Lüdtke, O., & Hall, N. C. (2011). Between-domain relations of academic emotions: Does having the same instructor make a difference?. *The Journal of Experimental Education*, *79*(1), 84-101.
Goetz, T., Lüdtke, O., Nett, U. E., Keller, M. M., & Lipnevich, A. A. (2013). Characteristics of teaching and students' emotions in the classroom: Investigating differences across domains. *Contemporary educational psychology*, *38*(4), 383-394.
Gunter, P. L., Shores, R. E., Jack, S. L., Rasmussen, S. K., & Flowers, J. (1995). On the move using teacher/student proximity to improve students' behavior. *Teaching Exceptional Children*, *28*(1), 12–14.
Hall, E. T., Birdwhistell, R. L., Bock, B., Bohannan, P., Diebold, A. R., Durbin, M., Edmonson, M. S., Fischer, J. L., Hymes, D., Kimball, S. T., La Barre, W., Frank Lynch, S. J., McClellan, J. E., Marshall, D. S., Milner, G. B., Sarles, H. B., Trager, G. L., & Vayda, A. P. (1968). Proxemics. *Current Anthropology*, *9*(2/3), 83–108.
Hektner, J., Schmidt, J., & Csikszentmihalyi, M. (2007). *Experience sampling method*. SAGE Publications, Inc.
Jennings, P. A., & Greenberg, M. T. (2009). The prosocial classroom: Teacher social and emotional competence in relation to student and classroom outcomes. *Review of Educational Research*, *79*(1), 491–525.
Kale, U. (2008). Levels of interaction and proximity: Content analysis of video-based classroom cases. *The Internet and Higher Education*, *11*(2), 119–128.
Lei, H., Cui, Y., & Chiu, M. M. (2018). The relationship between teacher support and students' academic emotions: A meta-analysis. *Frontiers in Psychology*, *8*.
Loderer, K., Pekrun, R., & Lester, J. C. (2020). Beyond cold technology: A systematic review and meta-analysis on emotions in technology-based learning environments. *Learning and instruction*, *70*, 101162.
Marshall, P. D., & Losonczy-Marshall, M. (2010). Classroom ecology: Relations between seating location, performance, and attendance. *Psychological Reports*, *107*(2), 567–577.
Mazer, J. P., McKenna-Buchanan, T. P., Quinlan, M. M., & Titsworth, S. (2014). The dark side of emotion in the classroom: Emotional processes as mediators of teacher communication behaviors and student negative emotions. *Communication Education*, *63*(3), 149-168.
McHugh, M. L. (2012). Interrater reliability: The Kappa statistic. *Biochemia Medica*, *22*(3), 276–282.
Meeks, M., Knotts, T., James, K., Williams, F., Vassar, J., & Wren, A. (2013). The impact of seating location and seating type on student performance. *Education Sciences*, *3*(4), 375–386.
Meyer D. K., & Turner J. C. (2007). Scaffolding emotions in classrooms. In Schutz P. A. & Pekrun R. (Eds.), *Emotion in education* (pp. 243–258). Elsevier.
Miller, P. W. (1988). *Nonverbal communication*. What research says to the teacher. In *ERIC* (3rd ed.). NEA Professional Library.
National Center for Education Statistics. (2023). *District Demographic Background 2017-2021*. ACS School

District Profile 2017-21.
Negoescu, A., & Boştină-Bratu, S. (2016). Teaching and learning foreign languages with ICT. *Scientific Bulletin*, *21*(1), 21–27.
Özdaş, F. (2022). Teachers' immediacy behaviors and academic achievement: A relational analysis. *SAGE Open*, *12*(1).
Pekrun, R., & Linnenbrink-Garcia, L. (2012). Academic emotions and student engagement. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 259–282). Springer Science + Business Media.
Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational Psychologist*, *37*(2), 91–105.
Pekrun, R., Lichtenfeld, S., Marsh, H. W., Murayama, K., & Goetz, T. (2017). Achievement emotions and academic performance: Longitudinal models of reciprocal effects. *Child development*, *88*(5), 1653-1670.
Pekrun, R. (2000). A social-cognitive, control-value theory of achievement emotions. In J. Heckhausen (Ed.), *Motivational psychology of human development: Developing motivation and motivating development* (pp. 143–163). Elsevier Science.
Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review*, *18*(4), 315–341.
Pekrun, R. (2024). Control-value theory: From achievement emotion to a general theory of human emotions. *Educational Psychology Review*, *36*, Article 83.
Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Rresearcher*, *38*(2), 109-119.
Plante, K., & Asselin, M. E. (2014). Best practices for creating social presence and caring behaviors online. *Nursing Education Perspectives*, *35*(4), 219–223.
Quinnell, R. (2022). *The impact of altering physical and human components in middle school mathematics classes on assessment performance*. Sophia, the St. Catherine University.
Rutherford, T., Liu, A. S., Lam, A. S., & Schenke, K. (2019). Impact on mathematics self-beliefs from a mastery-based mathematics software. *Journal of Research on Technology in Education*, *52*(1), 79-94.
Rutherford, T., Liu, A. S., & Wagemaker, M. (2021). "I Chose Math Because…": Cognitive interviews of a motivation measure. *Contemporary Educational Psychology*, *66*, 101992.
Sakiz, G., Pape, S. J., & Hoy, A. W. (2012). Does perceived teacher affective support matter for middle school students in mathematics classrooms?. *Journal of School Psychology*, *50*(2), 235-255.
Sánchez, L. M. (2014). *Emotions in classroom microsituations*: *A sociocultural perspective* [Doctoral dissertation, University of London Institute of Education]. University College London Discovery.
Savina, E., & Fulton, C. (2024). Students' and teachers' emotions in the classroom: An ecological dynamic systems perspective. *European Journal of Psychology of Education*, *39*, 2763–2781.
Schutz, P. A., & Lanehart, S. L. (2002). Introduction: Emotions in education. *Educational Psychologist*, *37*(2), 67–68.
Shores, R. E., Gunter, P. L., & Jack, S. L. (1993). Classroom management strategies: Are they setting events for coercion? *Behavioral Disorders*, *18*(2), 92–102.
Tan, J., Mao, J., Jiang, Y., & Gao, M. (2021). The influence of academic emotions on learning effects: A systematic review. *International Journal of Environmental Research and Public Health*, *18*(18), 9678.
Tze, V. M. C., Daniels, L. M., Klassen, R. M., & Li, J. C.-H. (2013). Canadian and Chinese university students' approaches to coping with academic boredom. *Learning and Individual Differences*, *23*, 32–43.
Vygotsky, L. S. (1978). *Mind in society: Development of higher psychological processes* (M. Cole, V. Jolm-Steiner, S. Scribner, & E. Souberman, Eds.). Harvard University Press.
Weaver, A. D., Scherer, P., Hengen, S., & Shriver, M. D. (2020). An exploratory investigation of proximity control in a large-group unstructured setting. *Preventing School Failure: Alternative Education for Children and Youth*, *64*(3), 261–270.
Weng, T.-S., Li, C.-K., & Hsu, M.-H. (2020). Development of robotic quiz games for self-regulated learning of primary school children. *Proceedings of 2020 AI and Cloud Computing Conference*, 58–62.

International Society of the Learning Sciences

Farm Harvests and Airplane Engines: Contextualizing Database Exercises in Relatable Contexts with a Learnersourcing Workflow

Yuzhe Zhou, Alejandra J Magana, Tianyi Li
zhou922@purdue.edu, admagana@purdue.edu, li4251@purdue.edu
Purdue University

Abstract: Developing learning materials that are relatable to students with varied backgrounds poses substantial challenges for instructors. This paper introduces a learnersourcing workflow designed to empower students as creators of personalized educational content, specifically through generating and peer-reviewing worked examples (WEs) that connect computing concepts to real-world scenarios. We empirically evaluated this learnersourcing approach within a database programming course and collected 127 student-generated WEs. The WEs contextualized database problems within diverse and novel scenarios, ranging from agricultural management to aircraft maintenance. Although some WEs had incomplete problem statements or incorrect solutions, most were complete, clear, and educational. In addition, student peer assessments of the WEs are aligned with instructor evaluations. We discuss the lessons learned regarding how a structured learnersourcing workflow can help students contextualize computing concepts and support learning.

Introduction

Learning materials that resonate with students' diverse backgrounds are crucial for building foundational programming knowledge (Bouvier et al., 2016; McKenna & Laycock, 2004). However, effective contextualization relies on contexts that are authentic and relatable to the learners (Griggs & Cox, 1982). Developing such learning materials requires substantial instructor effort and sensitivity to understand diverse student backgrounds.

We propose a learnersourcing (Kim, 2015) workflow that considers students as experts in their own backgrounds and interests, and experts-in-training in the computing concepts they are learning. The workflow guides students to create and peer-review worked examples (Renkl, 2002) as a channel to communicate how the computing concepts can be applied to address real-world problems relevant to them. The workflow was empirically evaluated in a database programming course with 45 junior and senior students. Students completed learnersourcing-based assignments covering different computing concepts.

While some student-generated examples occasionally had vague problem statements or incorrect solutions, most were clear, relevant, and educational within the course contexts. We identified 19 context themes from the student-created WEs and found close alignment between peer assessments and instructor evaluations.

This experience report contributes (1) a learnersourcing workflow and instructional design that enable students to contextualize computing concepts and collaboratively generate high-quality, relatable learning materials; and (2) empirical insights into the effectiveness of learnersourcing in producing student-centered worked examples and supporting student learning.

Related work

Relatable examples reduce student dropout rates (Guzdial, 2010), facilitate skill transfer (Bouvier et al., 2016), and support autonomous behavior aligned with self-determination theory (Assor et al., 2002; Patall et al., 2010). While the outcomes are beneficial, perception of relatability varies by prior knowledge (Wu, 2021), learning preferences (Smith & Renzulli, 1984), and cultural backgrounds (Fisher et al., 2012). This aligns with Culturally Responsive Teaching (CRT) (Scott et al., 2015), which values students' cultures and promotes inclusivity. However, contextualized learning materials by personal background require substantial instructor effort and sensitivity to diverse student experiences (Wyatt, 2015; Scott et al., 2015).

Learnersourcing in computing education

Learnersourcing, rooted in crowdsourcing and participatory learning (Kim, 2015), has the potential to address these challenges by transforming students from passive consumers to active contributors and co-creators of knowledge (Abdi et al., 2021; Singh, Brooks, & Shayan Doroudi, 2022). Notable implementations of learnersourcing in programming education include generating explanations of coding misconceptions (Guo et al., 2020), instructional hints (Glassman et al., 2016), and SQL exercises (Pirttinen et al., 2023). Platforms like

 © ISLS

PeerWise (Denny et al., 2008), RiPPLE (Khosravi et al., 2019), AXIS (Williams et al., 2016), and ExPeerience (Zhou et al., 2026) further facilitate the broader application of learnersourcing across various courses.

In line with the principles of culturally responsive teaching (Scott et al., 2015), we view students' diverse backgrounds as valuable assets and apply learnersourcing to guide students in co-creating relatable WEs. WEs are structured explanations that clearly present problem statements and detailed solution steps (Collins et al., 1991; Vieira et al., 2015).

Research testbed

We developed and refined our learnersourcing workflow within the context of an advanced database programming course offered at a large public university. The class covers PL/SQL (**P**rocedural **L**anguage extension for **SQL**), Oracle's procedural extension for Structured Query Language.

The majority of students enrolled in this course are junior or senior college students majoring in Information Technology or related fields. Enrollment in the course requires prior completion of an introductory SQL course. Class sizes range from 20-25 students in the spring semester to 40-50 in the fall semester, consistently maintaining a 1:4 female-to-male ratio that reflects the college's broader demographics. The preliminary study occurred in Spring 2022, and the evaluation study was conducted during Fall 2022.

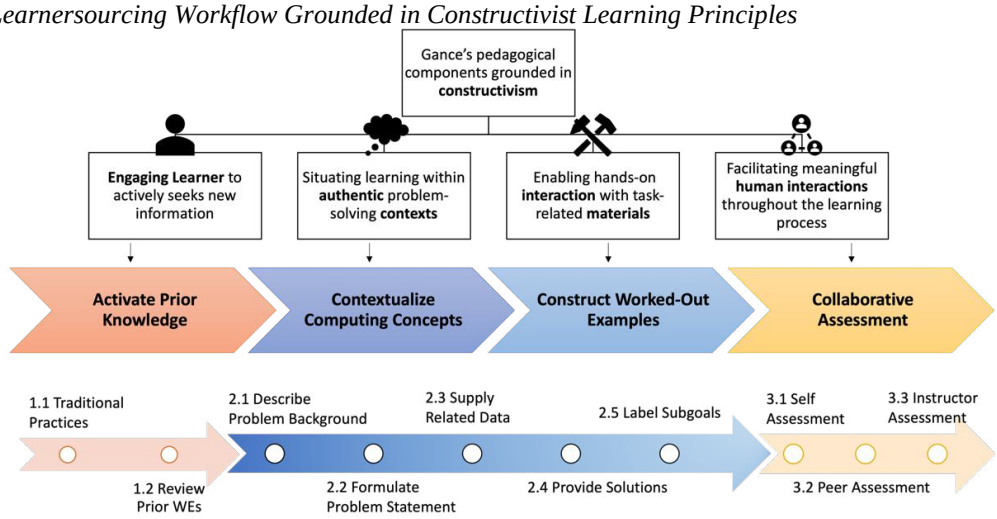
The course covers three main themes. *PL/SQL Fundamentals*: PL/SQL blocks, variables, conditional control structures, exception handling, and integration with SQL statements. *Functions and Procedures*: user-defined stored programs designed to encapsulate SQL and PL/SQL logic. *Triggers*: stored programs that execute automatically in response to defined database events.

Learnersourcing workflow design

We employed a design-based research approach structured around a three-phase design process. In Phase 1, we developed the learnersourcing workflow grounded in constructivist learning principles (Gance, 2002) (Figure 1, top). Phase 2 involved a preliminary study to identify key leverage points for scaffolding students. In Phase 3, we refined and evaluated the workflow design.

Figure 1

Learnersourcing Workflow Grounded in Constructivist Learning Principles



The final design of the learnersourcing workflow, delineating from the conceptual framework. The left red arrow at the bottom connects to the regular instruction activities. The middle gradient blue arrow guides students to contextualize computing concepts in their lived experience and construct worked examples. The right yellow arrow represents the collaborative assessment encompassing self, peer, and instructor assessment on the learnersourced worked examples.

Initial design and theoretical grounding

Our design of the learnersourcing workflow was guided by the constructivist framework proposed by Gance (2002), which emphasizes: (1) engaging learners, (2) enabling hands-on interaction with task-related materials,

(3) situating learning within authentic problem-solving contexts, and (4) facilitating meaningful human interactions throughout the learning process.

The initial workflow includes four components. First, *Activate Prior Knowledge* connects learners to existing course structures (e.g., lectures, labs). Second, *Contextualize Computing Concepts* prompts students to propose scenarios rooted in their personal interests. Third, *Construct Worked Examples* guides students to create step-by-step examples based on those scenarios. Finally, *Collaborative Assessment* fosters peer and self-evaluation, aligning with the constructivist emphasis on social learning.

Preliminary study and key leverage points

We implemented the workflow as bonus point opportunities within an online quiz and an assignment. Students were asked to design two practice questions, one easy and one medium, for classmates to practice *PL/SQL functions and procedures*. Students were required to outline relevant scenarios and database schemas and provide solutions with self-assessments. Each submitted WE underwent peer evaluations and independent instructor assessments using a rubric evaluating Difficulty, Clarity, Topic Relevance, and Educational Value. From a preliminary study, we collected 23 WEs and 100 peer reviews. Detailed elaboration and results are presented in another paper (Li et al., 2025).

Leveraging points for creating worked examples

The majority of students successfully created WEs relevant to course topics. Some problem statements did not provide a clear task regarding what should be done. Another challenge is the lack of *details* about the contexts, including the example *data*. In addition, many students did not come up with *new* contexts but rather reused or adapted the instructor's examples. The *quality of the solutions* created by students also varied by levels of detail and clarity. Comments from peer assessments also confirmed that the above-mentioned challenges would be confusing when engaging with WEs created by their peers.

Leveraging points and implications for peer assessment

We compared the self, peer, and instructor ratings using the Friedman test (Friedman, 1937). Results indicated statistically significant differences in Difficulty, Clarity, and Relevance ratings ($p < 0.001$). However, calculated effect sizes using Kendall's W were negligible ($w < 0.1$), suggesting the practical differences between groups' ratings were minimal despite statistical significance. Inter-rater agreements among peer reviewers were low, potentially due to varied interpretations of the evaluation criteria. Nonetheless, aggregating peer ratings across multiple reviewers provided sufficient reliability when compared with instructor ratings.

Revised workflow

Drawing on the lessons learned from the preliminary study, we improved the instructional design and assessment rubrics into the final design of the learnersourcing workflow (See Figure 1). The workflow is implemented within homework assignments.

Activate prior knowledge

The first task (pink arrow in Figure 1) is to answer two review questions, revisiting the key concept (*Step 1.1*). After that, a WE created by previous students in the preliminary study is presented as an example (with permission, *Step 1.2*). The opportunity to be featured in future classes also served as an extra source of motivation.

Contextualize computing concepts and construct worked examples

The second task (blue arrow in Figure 1) is to create the WE through five sub-steps.

The first sub-step is to identify a database problem from the students' lived experience, where PL/SQL can provide a solution. In this sub-step, students are asked to explain the problem context in a way that is understandable to someone unfamiliar with the topic (*Step 2.1*).

Next, students are tasked to define the problem formally and specify the expected answer (*Step 2.2*).

The third sub-step asks students to provide the data needed to solve the problem, with a minimum of two data tables, each with at least five attributes and 100 rows of dummy data (*Step 2.3*).

The final two sub-steps ask students to provide their solutions in a step-by-step manner, with at least three steps, each accompanied by a clearly labeled subgoal and explanation (*Steps 2.4 and 2.5*).

Collaborative assessment

The third task (yellow arrow in Figure 1) is to evaluate the student-generated worked examples. This task served both as a quality control mechanism and as an additional learning opportunity. By reviewing peers' worked examples, students engaged in evaluating different problem contexts and solution strategies, which can support

deeper conceptual understanding through explanation, comparison, and reflection. In addition, this process exposed students to how their peers approached the same programming concepts, allowing them to observe different problem-solving mindsets and ways of connecting abstract database concepts to real-world contexts.

Based on the pilot study results, we expanded the assessment rubrics to nine dimensions to define more specific and clear evaluation criteria (See Table 1). We also switched to a 7-step Likert scale, with 1 indicating Not at all, 7 for Very much, and 4 as neutral.

Table 1

Updated Assessment Metrics, The Corresponding Workflow Step Evaluated, and the Definition of Each Metric.

Criteria	Workflow Step	Definition
Context Usefulness	Step 2.1	The exercise shows how PL/SQL knowledge can be useful in a real-world scenario.
Problem Relevance	Step 2.1	The exercise requires applying PL/SQL functions and procedures to solve.
Problem Clarity	Step 2.2	The problem description was clear and easy to understand.
Data Quality	Step 2.3	Sufficient and meaningful example data were provided for this exercise.
Solution Clarity	Step 2.4 - 2.5	The step-by-step solution was clear and easy to understand.
Answer Quality	Step 2.4 - 2.5	The answers provided in each step were all correct.
Subgoal	Step 2.4 - 2.5	The steps provided by the exercise creator are necessary to solve the overall problem.
Difficulty Level	Overall	The exercise was too easy.
Educational Value	Overall	Working on the exercise develops a deeper understanding of the PL/SQL topic.

Workflow evaluation

We evaluated the revised learnersourcing workflow in the semester after the preliminary study.

Procedure

We deployed the learnersourcing workflow on all three course themes in the form of homework assignments. Homework 1 focused on PL/SQL blocks, Homework 2 focused on functions and procedures, and Homework 3 focused on triggers. Given that these homework assignments represent a novel learning activity still in its developmental stages, their grading was intentionally weighted relatively low (each homework is 4% of the final grade), with an emphasis placed on effort and engagement.

The homework assignments were released one week after the concepts were introduced in the lecture and practiced in the lab sessions. Students had two weeks to complete and submit each homework. Homework instructions are described in Table 2.

Table 2

Structure of The Homework Instruction and Corresponding Tasks

Part	Task	Description
A	Task 1.1	Answer two review questions related to the course topic.
B	Task 1.2	Review an example contextualized worked example (WE).
C	Task 2.1 – 2.5	Follow the steps shown in the example to construct their own WE on the topic.
D	Task 3.1	Self-assess the quality of the WE developed.

The homework has four parts (first column). The task numbers in the second column are aligned with the step numbers in Figure 1.

Students were asked to submit their answers to parts A and C of the instruction document within a single text file. The self-assessment (Part D) was administered through Circuit, an online peer-review platform developed at Purdue University that supports anonymous and asynchronous peer assessment, using 7-point Likert-scale questions for each metric listed in Table 1.

After the due date of submitting the WE, the Circuit tool randomly assigned three students to review each WE anonymously (*Task 3.2* in Figure 1). Each peer review includes the same questions in self-assessment, as well as an optional free-text comment.

Finally, the course instructor and teaching assistant reviewed every worked example created by students (*Task 3.3* in Figure 1) using the same scale and metric, hereby referred to as instructor ratings.

Data analysis

To identify themes in student-selected contexts, each author independently analyzed 10 randomly selected WEs from each homework. They then compared and synthesized their findings, resolving any disagreements. One author used the agreed-upon coding scheme to analyze the remaining WEs, and another reviewed a new random sample of 10 WEs. This process was repeated until no new disagreements arose.

To assess context integration, one author annotated justifications for the problem's necessity beyond a generic "given certain inputs, produce specified outputs." This involved noting whether the WEs mentioned reasons for solving the problem, such as addressing a business question or achieving a personal goal. Another author randomly selected 10 WEs from each homework to identify disagreements, repeating the process until no new disagreements arose.

Results overview

A total of 127 WEs were collected across three homework assignments. Out of the 127 WEs submitted, 115 received peer reviews. Additionally, many students provided optional free-text peer assessment comments. Table 3 summarizes the number of submissions received for each learning activity.

Table 3

Summary of Results

HW	WEs	Self	Peer	Comments
HW 1	42	42	110	23
HW 2	44	40	104	23
HW 3	41	33	80	17
Total	127	115	294	63

Summary of worked examples (WEs), self-assessments (Self), peer reviews (Peer), and free-text peer comments (Comments) submitted across homework assignments (HW)

Impact on the diversity, clarity, and integration of contexts

Overall, the revised learnersourcing workflow guided students to identify more diverse problem contexts that offered more in-depth knowledge about students' interests. We synthesized the contexts into 19 themes (See Table 4) with 13 from homework 1, 16 from homework 2, and 14 from homework 3.

Most contexts (107 out of 127) are notably different from the instructor's provided example. They incorporated elements tied to local events and businesses. Many students described the contexts in first- or second-person perspectives. Only four WEs had coding questions disconnected from the contexts or had missing contexts, where there was no logical connection between contexts and problems.

Similar to the preliminary study, video games and sports remained popular topics, but the computing concepts were better contextualized. For example, one student created a worked example about evaluating basketball player performance by analyzing how many seasons their favorite player averaged over 20 points per game (PPG) and shot at least 45% field goal (FG). He provided season-by-season data, including games played, total points, and number of shots made, and defined a performance evaluation metric. In contrast, although a worked example from the preliminary study also mentioned sports, it simply involved retrieving aggregated information from a single table of player statistics without further analysis or interpretation.

Impact on the quality of worked examples

Overall, instructor ratings of the WEs were positive across most quality metrics (See Figure 2). Over 95\% of the WEs from all three homework assignments received ratings above neutral (i.e., 5, 6, or 7) for nearly every quality metric, with the exceptions being difficulty and educational value.

The high instructor ratings for *context usefulness*, *problem relevance*, and *problem clarity* are aligned with the results described in the section above. Qualitative analysis revealed only one WE as irrelevant, as it omitted the required concept of PL/SQL triggers. Peer assessment comments further supported the positive instructor ratings regarding *data quality*, with only 20 instances of issues such as insufficient or missing data. This is also an improvement compared to the preliminary study, which had identified data quality problems in 11 out of 21 WEs.

Table 4

Distribution of Contexts Identified in WEs Across Homework Assignments

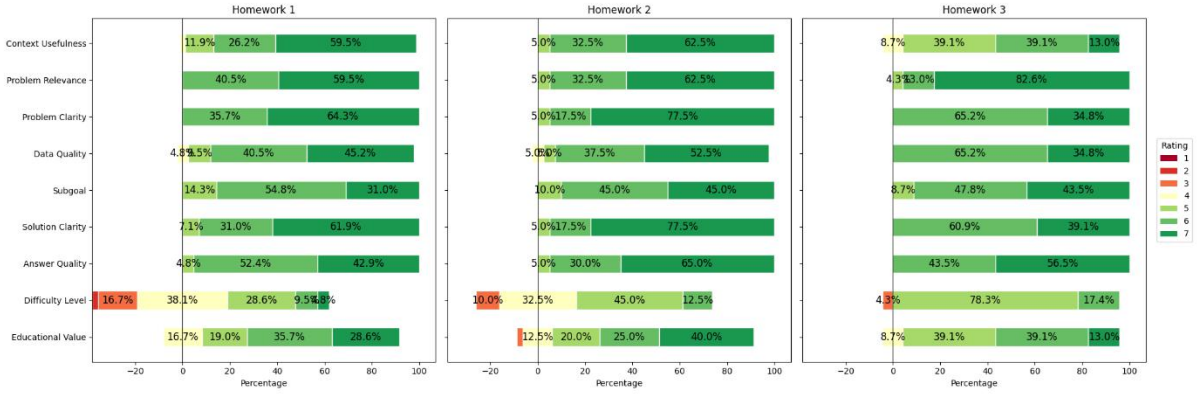
Themes	Description of Themes	HW1	HW2	HW3
Video Game	Video game selling, stories in video games	1	8	4
Sports	Player performance tracking, baseball tickets selling	4	2	0
Movie	Cinema people tracking, sales, customer tracking price, schedule	3	2	2
Students' information	Information related to students in university, such as majors, classes, and GPAs	1	2	3
Aerospace Business	Rocket booster business, airplane engine selling and repairing	2	0	0
Catering	Restaurant reservations, cake stores, and coffee shops sale tracking	4	2	3
Clothing	Clothes producing, selling, and orders	2	2	3
Tech Companies	Social media companies, 3D-printing technology, and IT Consulting	4	3	3
Online business	Online shopping platforms	5	2	5
Employee	Information related to employees such as income and hiring	5	2	3
Customer	Customer information such as phone numbers and zip codes	5	2	3
Shipping Services	Shipment tracking such as status of shipping and shipping date	4	3	6
Parts Selling	Business related to selling parts or technical equipment	2	5	3
Music	Types of music, and favorite music	0	2	0
Farms	Information related to farms such as harvest and equipment	0	3	1
Pet Care	Pet recording, Pet health tracking	0	1	0
HVAC Services	Heating and Air Conditioning Services.	0	1	0
Car Dealership	Cars selling, trading, and repairing	0	2	0
Pharmacy	Medicines and prescriptions	0	0	1

Regarding *subgoals*, most students structured their responses into four or five subgoals, exceeding the minimum requirement of three. However, some problems did not warrant multiple subgoals; as one peer commented, "...I don't think the solutions are evolved enough to have three steps." This sentiment is consistent with the instructors' *difficulty* ratings, with more than 80% of WEs being rated as neutral or somewhat easy.

Furthermore, the difficulty and educational value of WEs improved over time. In Homework 1, more than half of the WEs were rated as easy to neutral (ratings of 1–4), while this percentage dropped to less than half in Homework 2, and only 4.3% in Homework 3. For educational value, 83.3% of the WEs in Homework 1 were rated above neutral, compared to 85% in Homework 2 and 92.3% in Homework 3 (See Figure 2).

Figure 2

Distribution of Instructor Ratings for Worked Examples across Three Homework



Ratings were provided on a 7-point Likert scale (1 = Not at all, 4 = Neutral, and 7 = Very much so). The numbers indicate the percentage of student submissions within each homework assignment that received each rating.

 © ISLS

Most WEs received ratings in the *answer quality* range of 5, 6, or 7. Specifically, submissions that compiled, executed correctly, and fully solved the problem were awarded a 7, while error-free answers that deviated slightly from the expected steps received a 6. Answers with minor compilation or execution issues were generally rated between 4 and 5, with non-executable code receiving a rating of 1.

Reliability of student self and peer assessment

We conducted a Friedman test to compare the ratings provided by self-assessment, peer assessment, and instructor assessment. The results indicated statistically significant differences among the three types of ratings ($p < 0.05$), although the effect sizes were negligible ($w < 0.1$; see Table 5).

Table 5
Results of Friedman Tests on The Rating Similarity for Each Criterion.

Evaluation Criteria	Chi^2	df	p-value	Kendall's W
Context Usefulness	12.82	2	0.002	0.06
Clarity of Problem	22.41	2	<.001	0.11
Clarity of Solution	14.63	2	0.001	0.07
Data Quality	9.49	2	0.009	0.05
Answer Quality	6.63	2	0.036	0.03
Difficulty	26.42	2	<.001	0.13
Subgoal	9.72	2	0.008	0.05
Relevance	16.18	2	<.001	0.08
Educational Value	26.72	2	<.001	0.13

We then analyzed the inter-rater agreement among peer reviewers using Krippendorff's alpha. Most (87 out of 105) WEs have limited agreement (alpha < 0.667) among the peer reviews. Combining the quantitative results with the qualitative insights from the free-text comments, it appears that the low rates of agreement are primarily attributable to the varying levels of effort exerted in the peer assessments. Specifically, some peer reviewers pointed out issues, while others focused on efforts. In Homework 1, out of the 20 WEs that received two or three free-text comments, 10 demonstrated contrasting feedback. Homework 2, the number is eight out of 21 WEs, and in Homework 3, it is eight out of 16.

Discussion

In this work, we designed and evaluated a learnersourcing workflow to involve students in co-creating relatable learning materials. This work opens up exciting possibilities for leveraging the learnersourcing approach in the empirical classroom settings, with implications for both teaching and system design to support learners.

Eliciting diverse and relatable contexts from students

The revised learnersourcing workflow guided the students to produce WEs that are not only contextualized but also highly relevant to the specified course topic. The workflow successfully elicited more varied contexts, moving beyond the instructor's examples and reflecting diverse student interests. The majority of the WEs from the final evaluation were reasonably contextualized and meaningfully integrated the context with the computing concepts.

This study introduces a compelling avenue for instructors navigating the intricate task of tailoring learning materials for diverse students, which often requires an in-depth understanding of each student's unique background (Trinidad, 2020; Brown, 2003). Leveraging the collective intelligence and diverse perspectives of students can contribute to enriching the variety of contexts within learning materials. Over time, this process can build a repository of contextually rich examples that instructors can curate and reuse in future course offerings. Such repositories may help instructors continuously update course materials while maintaining relevance to students' evolving interests.

Learnersourcing as an artifact to implement social learning theories

Grounded in constructivist principles, the learnersourcing workflow transforms students from passive recipients to active agents in knowledge creation, merging computational concepts with their real-life experiences. The WEs demonstrate this blend, with students applying theoretical knowledge to varied real-world contexts. In addition,

 © ISLS

our results resonate with the principles of the protégé effect (Chase et al., 2009; Roscoe & Chi, 2007), which posits that students deepen their own understanding and skills by teaching their peers. This is manifested in the high-quality WEs as well as the active participation and enthusiastic peer assessment comments.

The learnersourcing approach also has the potential to shift the classroom dynamics. Empowering students to contribute content and insights alters the traditional power dynamic, positioning students not just as learners but as active participants and contributors in the educational process. For instructors, the contexts created by students, along with any errors in the creation and assessment phases, provide deep insights into students' learning progress and interests. Student-created materials provide indirect feedback on instructional efficacy, suggesting areas where students may struggle or lack interest, thus informing necessary pedagogical refinements.

Although this study focuses on database programming, the learnersourcing workflow is not specific to PL/SQL. The core design principles, eliciting students' lived experiences, structuring the creation of worked examples, and integrating peer evaluation, can be adapted to other computing domains that involve problem solving and conceptual transfer. Instructors can modify the scaffolding prompts and evaluation criteria to align with different technical topics while preserving the overall workflow structure.

Associating educational values to quality control of learnersourced materials

Our preliminary study reveals several quality control challenges, and the revised workflow demonstrated effective mitigation of those challenges. Similar to crowdsourcing, quality control of learnersourced materials in our studies primarily relied on scaffolding instructions and self-assessment. Step-by-step guidance to describe the problem context and develop worked examples effectively assisted students in contemplating and explaining their contexts. The nine reviewing criteria served as additional guidance for students to understand the learnersourcing tasks and expectations on the contextualized worked examples to create.

In addition, the self and peer assessment can be regarded as an additional learnersourcing process embedded in the workflow. As students worked on and reviewed their peers' WEs, they not only practiced applying computational concepts in various real-world contexts but also gained insights into how their classmates structured solutions, thus broadening their understanding of applying computing concepts to new situations. Our learnersourcing workflow has great potential to be combined with prior successful learnersourcing methods (Glassman et al., 2016; Guo et al., 2020) to elicit richer peer feedback, such as hints and debugging suggestions on the WE solutions.

Limitations and future works

Our evaluation assessed the effectiveness of a learnersourcing workflow in eliciting diverse and relatable contexts from students. While promising, future research is needed to examine students' perceptions, workload concerns (Kyndt & Lydwin Bulckens, 2014), and engagement, as well as use controlled studies to formally quantify learning outcomes. Given variations in students' cultural, social, and academic backgrounds, long-term evaluations are needed to understand how these demographic factors influence the workflow's effectiveness across different course levels and subjects. Finally, despite students' valuable contributions, not all WEs are ready for direct integration into course materials. More rigorous quality assurance to address errors, formatting, and solution accuracy is required before student-generated WEs become standard course resources.

Conclusion

This paper introduces and evaluates a learnersourcing workflow that empowers students to create Worked Examples (WEs) based on their experiences and interests. Our evaluation showed that students produced high-quality, relatable learning materials across various real-world contexts. The learnersourcing approach leverages students' inherent expertise in their own backgrounds and interests to co-create relatable learning materials. In addition, peer assessment can serve as a reliable quality assurance mechanism for learnersourced materials. While most WEs were clear and valuable, maintaining consistent quality is important. This work highlights learnersourcing's potential to enrich educational content and support personalized, inclusive learning while providing instructors with valuable insights into students' progress.

References

Renkl, A. (2002). Worked-out examples: Instructional explanations support learning by self-explanations. *Learning and instruction*, *12*(5), 529-556.

Abdi, S., Khosravi, H., Sadiq, S., & Demartini, G. (2021). Evaluating the quality of learning resources: A learnersourcing approach. *IEEE Transactions on Learning Technologies, 14*, 1. https://doi.org/10.1109/TLT.2021.3058644

Assor, A., Kaplan, H., & Roth, G. (2002). Choice is good, but relevance is excellent: Autonomy-enhancing and suppressing teacher behaviors predicting students' engagement in schoolwork. *The British Journal of Educational Psychology*, *72*, 261–278. https://doi.org/10.1348/000709902158883

Bouvier, D., Lovellette, E., Matta, J., Alshaigy, B., Becker, B. A., Craig, M., Jackova, J., McCartney, R., Sanders, K., & Zarb, M. (2016). *Novice programmers and the problem description effect*. 103–118. https://doi.org/10.1145/3024906.3024912

Brown, K. L. (2003). From teacher-centered to learner-centered curriculum: Improving learning in diverse classrooms. *Education*, *124*, 1.

Chase, C. C., Chin, D. B., Oppezzo, Marily A, & Schwartz, D. L. (2009). Teachable agents and the protégé effect: Increasing the effort towards learning. *Journal of Science Education and Technology*, *18*, 334–352.

Collins, A., Brown, J. S., & Holum, A. (1991). Cognitive apprenticeship: Making thinking visible. *American Educator*, *15*, 3.

Denny, P., Hamer, J., Luxton-Reilly, A., & Purchase, H. (2008). PeerWise: Students sharing their multiple choice questions. *Proceedings of the Fourth International Workshop on Computing Education Research*, 51–58. https://doi.org/10.1145/1404520.1404526

Fisher, D., Frey, N., & Lapp, D. (2012). Building and activating students' background knowledge: It's what they already know that counts: Teachers must assess and build on the background knowledge students possess. *Middle School Journal*, *43*, 3.

Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association*, *32*, 200. https://doi.org/10.1080/01621459.1937.10503522

Gance, S. (2002). *Are constructivism and computer-based learning environments incompatible?* Ann Arbor, MI: MPublishing, University of Michigan Library.

Glassman, E. L., Lin, A. E., Cai, C. J., & Miller, R. F. (2016). Learnersourcing Personalized Hints. *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. https://doi.org/10.1145/2818048.2820011

Griggs, R. A., & Cox, J. R. (1982). The elusive thematic-materials effect in Wason's selection task. *British Journal of Psychology*, *73*, 3. https://doi.org/10.1111/j.2044-8295.1982.tb01823.x

Guo, P. J., Markel, J. M., & Zhang, X. (2020). *Learnersourcing at scale to overcome expert blind spots for introductory programming: A three-year deployment study on the python tutor website*. *Proceedings of the Seventh ACM Conference on Learning@ Scale*. 301–304. https://doi.org/10.1145/3386527.3406733

Guzdial, M. (2010). Does contextualized computing education help? *ACM Inroads*, *1*, 4. https://doi.org/10.1145/1869746.1869747

Khosravi, H., Kitto, K., & Williams, J. (2019). RiPPLE: A crowdsourced adaptive platform for recommendation of learning activities. *Journal of Learning Analytics*, *6*. https://doi.org/10.18608/jla.2019.63.12

Kim, J. (2015). *Learnersourcing: improving learning with collective learner activity*. (Doctoral dissertation, Massachusetts Institute of Technology).

Kyndt, E., & Lydwin Bulckens. (2014). "Time is not enough."Workload in higher education: a student perspective. *Higher Education Research & Development*, *33*, 4.

Li, T., Zhou, Y., & Magana, A. J. (2025, June). Towards Learnersourcing Relatable and Contextualized Learning Materials: An Exploratory Study in a Database Programming Class. In *Proceedings of the 30th ACM Conference on Innovation and Technology in Computer Science Education V. 1* (pp. 479-485).

McKenna, P., & Laycock, B. (2004). Constructivist or instructivist: pedagogical concepts practically applied to a computer learning environment. *Proceedings of the 9th Annual SIGCSE Conference on Innovation and Technology in Computer Science Education - ITiCSE '04*. https://doi.org/10.1145/1007996.1008041

Patall, E. A., Cooper, H., & Wynn, S. R. (2010). The effectiveness and relative importance of choice in the classroom. *Journal of Educational Psychology*, *102*, 4.

Pirttinen, N., Hellas, A., & Leinonen, J. (2023). Experiences from learnersourcing SQL exercises: Do they cover course topics and do students use them? *Proceedings of the 25th Australasian Computing Education Conference*, 123–131. https://doi.org/10.1145/3576123.3576137

Roscoe, R. D., & Chi, M. T. (2007). Understanding tutor learning: Knowledge-building and knowledge-telling in peer tutors' explanations and questions. *Review of Educational Research*, *77*, 4.

Scott, K. A., Sheridan, K. M., & Clark, K. (2015). Culturally responsive computing: A theory revisited. *Learning, Media and Technology*, *40*, 4.

Singh, A., Brooks, C., & Shayan Doroudi. (2022). Learnersourcing in Theory and Practice: Synthesizing the Literature and Charting the Future. *Proceedings of the Ninth ACM Conference on Learning@ Scale*. https://doi.org/10.1145/3491140.3528277

Smith, L. H., & Renzulli, J. S. (1984). Learning style preferences: A practical approach for classroom teachers. *Theory into Practice*, *23*, 1.

Trinidad, J. E. (2020). Understanding student-centred learning in higher education: students' and teachers' perceptions, challenges, and cognitive gaps. *Journal of Further and Higher Education*, *44*, 8. https://doi.org/10.1080/0309877X.2019.1636214

Vieira, C., Yan, J., & Magana, A. J. (2015). Exploring design characteristics of worked examples to support programming and algorithm design. *The Journal of Computational Science Education*, *6*(1), 2–15. https://doi.org/10.22369/issn.2153-4136/6/1/1

Williams, J. J., Kim, J., Rafferty, A., Maldonado, S., Gajos, Krzysztof Z, Lasecki, W. S., & Heffernan, N. (2016). AXIS: Generating explanations at scale with learnersourcing and machine learning. *Proceedings of the Third (2016) ACM Conference on Learning @ Scale*, 379–388. https://doi.org/10.1145/2876034.2876042

Wu, W. (2021). SQL2X. *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*. https://doi.org/10.1145/3408877.3432541

Wyatt, T. (2015). Understanding the process of contextualization. *Multicultural Learning and Teaching*, *10*

Zhou, Y., Manjunatha Babu, P., Pandey, U., Magana, A. J., & Li, T. (2026). ExPeerience: Towards AI-assisted learnersourcing to bridge conceptual understanding and problem solving in database programming education. *Proceedings of the 31st International Conference on Intelligent User Interfaces (IUI '26)*, 1178–1201. https://doi.org/10.1145/3742413.3789139

Acknowledgments

This work is supported by the National Science Foundation (NSF) under grant numbers 2315683 and 2434429. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

International Society of the Learning Sciences

Bridging Curriculum and Creativity: NLP-Powered Semantic Analysis for Knowledge Building

Ahmad Khanlari, University of Toronto, a.khanlari@mail.utoronto.ca

Abstract: Knowledge Building is a collaborative approach that treats ideas as improvable knowledge artifacts and engages learners in advancing collective understanding. To support this, educators need analytics that reveal deep discourse patterns. This paper introduces a semantic analysis tool for Knowledge Building that uses natural language processing to illuminate conceptual dynamics in student-generated content. The tool takes a glossary of curriculum terms and student notes, preprocesses text with spaCy, and computes semantic similarities using a transformer model. By analyzing word-to-term relationships, it identifies curriculum coverage, related concepts beyond official terms, and evidence of conceptual growth. We outline the tool's design, demonstrate its outputs through network visualizations, and discuss implications for formative assessment, curriculum design, and longitudinal tracking, positioning it as a flexible analytic for studying and fostering knowledge-creating discourse.

Introduction

Knowledge Building (KB) is an educational theory and pedagogy developed by Scardamalia and Bereiter (2006) that emphasizes the creation and continual improvement of ideas within a community of knowledge builders. In a KB community, students collectively take responsibility for advancing shared knowledge—treating ideas as communal objects to refine, rather than simply as personal opinions (Scardamalia & Bereiter, 2006). This approach, going beyond traditional knowledge acquisition, emphasizes on engaging learners in practices akin to those of scientific and creative communities (e.g. theory-building, questioning, explaining) (Scardamalia & Bereiter, 2014). Over decades of research, twelve KB principles (e.g. *Community Knowledge, Idea Improvement, Epistemic Agency, Embedded and Transformative Assessment*) have been articulated to guide classroom practice (Scardamalia, 2002). These principles underscore that students should not only learn content but also engage in the higher-order process of creating new knowledge and assessing their collective progress. For example, *embedded and transformative assessment* calls for assessment that is woven into the learning activities, helping the community reflect and improve in real time, rather than external, summative evaluation. Such theoretical foundations highlight the need for tools that make the often invisible dynamics of idea development visible and actionable to learners and educators.

Central to Knowledge Building is Knowledge Forum (KF), an online environment purpose-built to support knowledge creation through engaging students in asynchronous discourse where they post "notes" to share ideas, build on peers' contributions, and co-develop theories (Scardamalia, 2004). Knowledge Forum itself has an evolving suite of built-in analytic tools aimed at *"making thinking visible"* for participants (Burtis, 1998; Zhang et al., 2007). These include basic measures of online activity (e.g. notes written, read, and built-upon) as well as more content-oriented tools. However, standard Knowledge Forum analytics (e.g. post counts, reply counts, word cloud) offer only *very limited insight* into the quality of discourse. Teachers cannot easily discern, for instance, which key concepts have been grasped or how creatively students are connecting ideas, simply from quantity-based metrics. Prior research in Computer-Supported Collaborative Learning (CSCL) and Knowledge Building communities has therefore stressed the need for more advanced analytics attuned to discourse quality and knowledge dynamics (e.g., Jung et al., 2025). In line with this need, a variety of analytic tools have been developed to support embedded assessment in Knowledge Building environments (Zhu & Kim, 2017). Several analytic and visualization tools have contributed to this goal. However, each existing tool has specific focuses and limitations. For example, *Comparative Word clouds* (Resendes et al., 2015) provide quick overviews of frequently used terms and can highlight alignment between student and expert vocabulary, but they rely on exact matches and ignore context or synonyms. *KBDeX* (Oshima et al., 2012) is another analysis tool that uses socio-semantic network analysis to visualize collaboration patterns and conceptual clusters but connects words only by co-occurrence rather than meaning. Similarly, *Vocabulary Analyzer* and *Lexical Analyzer* (Hong et al., 2015; Laferrière et al., 2016) track vocabulary growth using curriculum-based lists, offering valuable but limited insights since they miss emergent or rephrased ideas.

Overall, none of the tools directly measures *semantic similarity* or the presence of *conceptually related terms* that are not exact matches. In a KB context, where students are encouraged to use their own words and even coin new terms or analogies, this is a crucial gap. Our work responds to this call by introducing a semantic analysis

 © ISLS

tool that can flexibly detect conceptual connections. By leveraging modern Natural Language Processing (NLP) techniques, the tool complements existing analytics: it goes beyond counting specific words to identifying when different words *mean* similar things, thereby revealing latent thematic threads and conceptual progress in the discourse.

Description of the semantic tool

The proposed semantic analysis tool is designed to examine *word-to-term similarities* within Knowledge Building discourse. By focusing on words and their semantic relationships, the tool captures the subtle ways students express key ideas—even when they do not use the exact curriculum terms. This makes it possible to recognize both *direct alignment* with disciplinary vocabulary and *creative extensions* that signal the emergence of new ideas. Indeed, this tool goes beyond identifying exact matches: it uses a transformer-based model to compute semantic similarity scores, revealing how closely student language relates to glossary terms and how alternative expressions may still connect to core concepts. In this way, the analysis provides insight into both the *depth of curriculum engagement* and the *diversity of student contributions* that drive collective knowledge advancement.

In the sections that follow, we describe the tool's inputs, the natural language processing pipeline, the semantic similarity computations via transformers, and the overall conceptual architecture that supports these analyses.

Data input

The tool begins with two inputs: a glossary of terms and the set of student notes written in Knowledge Forum. The glossary represents the *target conceptual space*—for example, a set of curriculum keywords, scientific concepts, teacher-defined "big ideas" for the unit, or a collaboratively built class glossary. The student notes are the raw texts contributed by students over some period, representing the community's collective discourse. Each note typically contains a title and a body of text written by a student, possibly with quotes from other notes (build-ons) and scaffold supports.

Preprocessing with spaCy

Both datasets are cleaned and prepared using *spaCy* (Honnibal, 2017), a widely used language analysis toolkit for cleaning and lemmatization. During the lemmatization step, words are normalized into their base forms (e.g., "recycling" → "recycle", "trees" → "tree") so that different grammatical variations are counted together. This step is crucial because students may use different grammatical forms of the same word. Lemmatization reduces words to their base form, ensuring that our analysis counts them as the same concept when appropriate.

Semantic modeling with transformers

At the core of the tool's capability is its ability to compute semantic similarity scores between words/terms in student notes and terms in the glossary. Rather than relying on simple keyword or root matches, the tool employs a modern *transformer-based model* (Wang et. al., 2017) named "*all-MiniLM*" that captures deeper conceptual relationships.

Transformer-based models represent a major advance in natural language processing, enabling systems to capture the contextual meaning of words and phrases rather than treating them as isolated tokens. At the core of transformers is the *self-attention mechanism* (Vaswani, et al., 2017), which allows the model to weigh the importance of each word in a sequence relative to others. This means that the representation of a word like *cell* will differ depending on whether it appears in a biology or prison context. By stacking multiple layers of attention and feed-forward networks, transformers learn rich, high-dimensional embeddings that capture semantic relationships between terms.

The tool described in this paper employs *all-MiniLM*, a lightweight but powerful transformer from the Sentence-BERT family (Reimers & Gurevych, 2019). MiniLM achieves strong performance in semantic similarity tasks while being computationally efficient. It maps words, phrases, or sentences into dense vector spaces where semantically related terms are placed close together. For example, *"habitat"* and *"home"* appear near each other in the embedding space, even though they are different words. This property makes all-MiniLM particularly suited for analyzing student discourse in Knowledge Building: it enables the detection of conceptual alignment with glossary terms even when students use paraphrases, synonyms, or creative expressions. In this way, transformer-based embeddings go beyond exact word matching to reveal the deeper conceptual dynamics of discourse.

Leveraging transformers will enable the tool to not only identify when students use glossary terms directly but also detect when they are discussing *related or conceptually adjacent ideas*. For example, even if a student does not use the glossary term *photosynthesis* directly, the tool can still capture their engagement with the

concept. If their notes include related terms such as *chlorophyll* or *carbon dioxide*, these words will appear close to *photosynthesis* in the embedding space. This proximity reveals that the student is implicitly addressing the glossary concept, even without explicitly naming it. From an educational perspective, this semantic approach makes visible both *curriculum-aligned understanding* and *emergent connections beyond the glossary*, helping instructors and students see how discourse reflects not only formal knowledge but also the *creative and exploratory aspects of knowledge building.*

Similarity computation

The tool compares every glossary term with student words and terms in two ways:

1. **Exact Matches** – If a student uses a glossary term directly, the connection is marked as a perfect match.
2. **Semantic Matches** – If students use their own words that are close in meaning to glossary terms (e.g., "bugs" instead of "insects"), the tool computes a similarity score indicating how close these words are conceptually.

Cosine similarity is a standard similarity score that ranges from -1 to 1, where 1 means the vectors point in the exact same direction (identical meaning in this context) and 0 means orthogonal (unrelated). For example, suppose one glossary term is *"habitat"* and a student note contains the word *"home"* in the context of animals. The transformer model might place *"habitat"* and *"home"* reasonably close, yielding a similarity score -say- around 0.7.

Filtering and visualization

To avoid overwhelming detail, the tool filters results using a similarity threshold so that only the most meaningful matches are displayed. Crucially, this threshold is customizable, allowing educators to adapt the analysis to different knowledge building goals. A higher threshold foregrounds only the strongest concept-to-concept links, helping instructors and students identify the core ideas that anchor collective inquiry. It also highlights the extent to which students are working within curriculum expectations, offering a focused view of how discourse aligns with established learning goals. By contrast, a lower threshold opens the analysis to a broader range of connections, capturing tentative, exploratory, and peripheral ideas that may not yet align tightly with the curriculum language. From a knowledge building perspective, this looser view is valuable because it makes visible the creative extensions, emergent concepts, diverse ways of expressing understanding, and even ideas that go beyond curriculum expectations—all of which can fuel deeper discourse and collective progress. In this way, adjusting the threshold functions as a pedagogical choice: "zooming in" to highlight consolidation, or "zooming out" to encourage exploration and idea diversity.

The retained connections are displayed in a network graph with an intuitive color system:

- 🔴 **Red nodes** represent glossary terms (i.e., curriculum concepts) that were not explicitly used by students in their written notes. These terms are present in the curriculum or conceptual framework provided by the instructor but do not appear verbatim in the students' discourse. Red nodes help identify *potential conceptual gaps* in students' engagement with core ideas.
- 🟢 **Green nodes** represent glossary terms that students explicitly used in their notes. This indicates direct alignment between the curriculum and student contributions. The shift from red to green occurs dynamically in the graph when a glossary term is found in the student-generated content. These green nodes make it easy to visually track *coverage of key curriculum terms.*
- 🔵 **Blue nodes** represent student-generated words that were not in the glossary but are semantically similar to one or more glossary terms, with a similarity score above a given threshold. These reflect students' own language, paraphrasing, or expansion beyond the provided curriculum terms. Blue nodes are particularly valuable for identifying *emergent ideas*, *student agency*, and *conceptual extensions* that go beyond the original terminology.

This color-coding approach allows researchers and educators to easily identify not only which curriculum concepts were addressed directly, but also how students may have rephrased or extended those concepts using their own vocabulary. The visualization thus provides insight into both *curriculum coverage* and *conceptual innovation* within a Knowledge Building environment.

Edge color coding based on semantic nearness

In addition to node colors, the graph employs color-coded edges to represent the strength of semantic similarity between glossary concepts and student-generated terms. Each edge connects a glossary term to a word or phrase from the student notes that the transformer model identifies as semantically related.

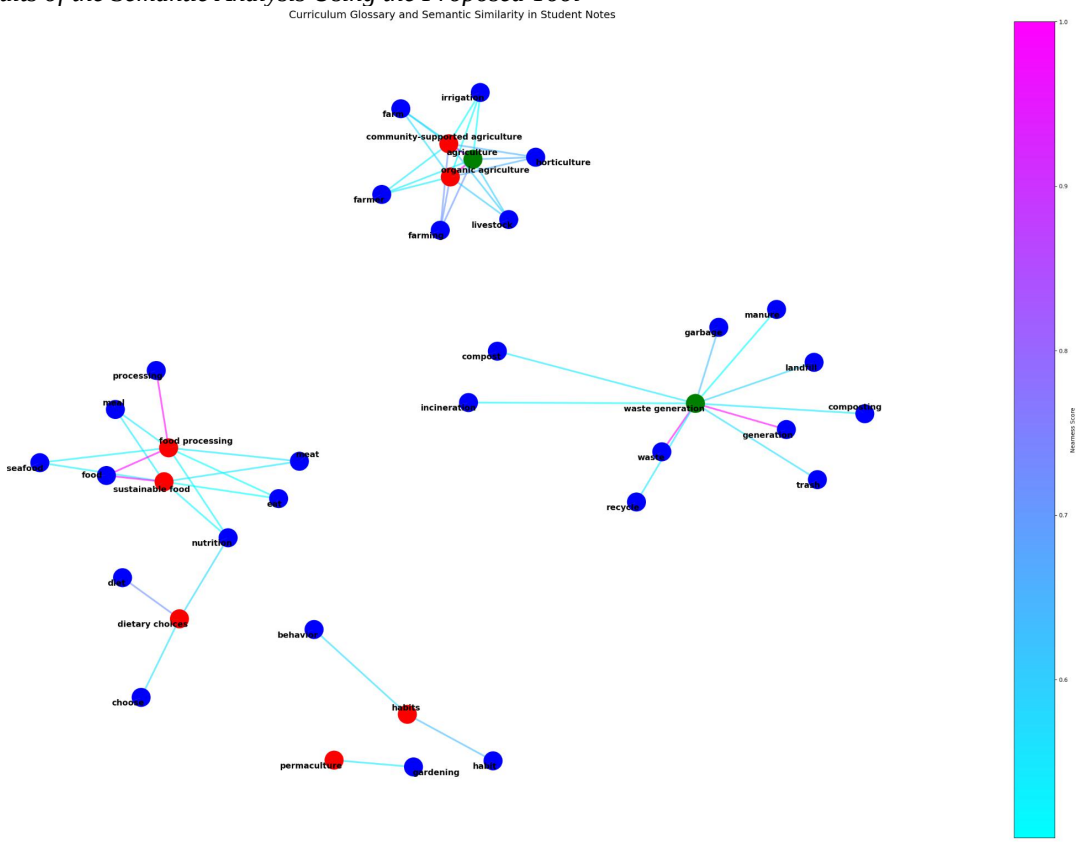
The *intensity and hue of each edge color* correspond directly to the similarity score, as indicated by the legend on the right. Lighter cyan edges represent lower similarity values (closer to the defined threshold), while deeper magenta edges represent stronger similarities (approaching 1.0). By consulting the legend, viewers can translate any edge color into its approximate similarity score, making it possible to distinguish strong conceptual alignments from weaker, more tentative connections. Figure 1 presents the outcomes of analyzing a view of student notes using this innovative tool, offering a detailed visualization that reveals the semantic relationships, conceptual alignments, and potential knowledge gaps within the Knowledge Forum discourse.
This design allows the visualization to go beyond a simple binary match/no-match view: it presents a *semantic gradient* that highlights not only whether a student's word relates to a glossary concept, but also how strongly it does so:

- A nearness score approaching 1 (indicating very high similarity) will be represented by an edge in a dark, vivid shade, signaling strong alignment with a glossary concept.
- An edge with a score close to the threshold (moderate similarity) will appear in a lighter hue, reflecting a looser, more exploratory link.

From a Knowledge Building perspective, darker edges indicate where student discourse remains tightly anchored to curriculum expectations, while lighter edges reveal emergent or exploratory ideas that may signal creative extensions, tentative reasoning, or steps toward moving beyond the formal curriculum. Together, these gradients provide educators and researchers with a nuanced picture of student thinking—ranging from the consolidation of established concepts to the exploration of new conceptual territory. This allows for a more precise assessment of how closely student discourse aligns with expected curriculum while also illuminating moments where students push the boundaries of the curriculum to generate novel insights.

Figure 1

Results of the Semantic Analysis Using the Proposed Tool

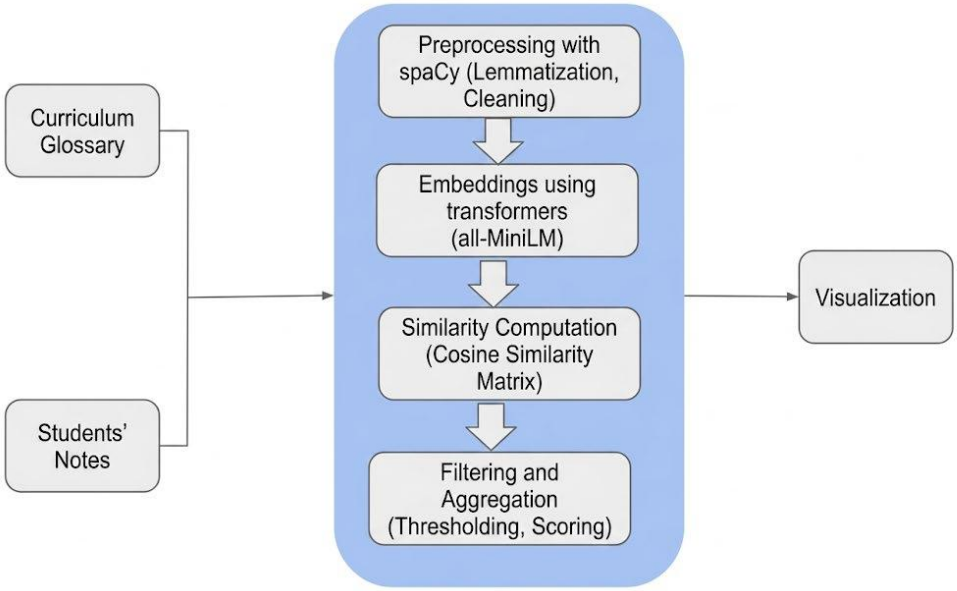


As illustrated in Figure 1, the visualization—generated by applying the analytics tool to a small dataset—reveals the extent to which students incorporate glossary terms related to waste generation and agriculture in their Knowledge Forum notes. Key glossary terms such as "waste generation" and "agriculture," highlighted in green within the visualization, indicate their active usage by students in their Knowledge Forum notes, demonstrating alignment with the provided authoritative glossary. However, several key glossary terms, marked in red, are absent

from student discourse, including "food processing," "sustainable food," and "organic agriculture." Despite these gaps, students introduced related terms (blue nodes) such as "meat," "nutrition," "seafood," "processing," "farming," and "irrigation," suggesting engagement with conceptually adjacent ideas. The edge colors in the graph denote semantic similarity, with a notable example being the term "food processing" from the glossary and "processing" in student notes, connected by a pink edge indicating a high similarity score of approximately 0.9. This suggests that while students may not use the exact glossary terms, they are exploring closely related concepts, demonstrating partial conceptual alignment.

In summary, the tool's architecture can be understood as a pipeline, which is shown in Figure 2.

Figure 2
Block Diagram of the Semantic Analysis Tool Workflow



Having described how the tool operates behind the scenes, we now turn to its novel capabilities for analyzing and interpreting Knowledge Building discourse.

Novel capabilities of the semantic tool

By applying the above methodology, our tool unlocks several novel analytic capabilities that extend beyond what prior Knowledge Building tools offer. In this section, we discuss these capabilities in turn, explaining how each addresses a specific need in Knowledge Building analytics. The key affordances include: (1) gauging *curriculum coverage* in discourse, (2) identifying the *use of related concepts beyond the given curriculum*, and (3) illuminating *cross-topic bridges* in discourse. We describe each capability and provide examples of the kinds of insights the tool can reveal, emphasizing how they can inform teachers' and researchers' understanding of the community's knowledge progress.

Curriculum coverage via semantic matching

The most immediate use of the tool is to determine to what extent students have covered the intended curriculum concepts in their discussions. Traditional keyword checks only flag when specific terms are mentioned; our tool goes further by flagging when concepts are discussed in *any terms*. For each term in the teacher's glossary, the tool finds any semantically related words that students used. This means that even if a concept's official name was never mentioned, we can recognize its *presence* in the conversation. For instance, suppose *"photosynthesis"* is a key term in a biology class. If students never wrote "photosynthesis" explicitly, a simple keyword search would wrongly conclude they ignored it. But our tool might find that students frequently discussed plants producing food with sunlight, using words like "energy", "sunlight", "make food", "chlorophyll". These words would show high similarity to "photosynthesis", and thus the tool would indicate that the concept was indeed covered implicitly. In a real scenario, one could imagine the output:

- **Term:** Photosynthesis

- **Related student words: sunlight** (similarity score: 0.7), **food** (similarity score: 0.75), **energy** (similarity score: 0.75), **chlorophyll** (similarity score: 0.85)

Such a report tells the teacher that students are engaging with aspects of photosynthesis, even if they never use the exact term. In this case, the teacher might conclude that students have developed a working understanding of the concept and simply need to connect it to the formal scientific terms. Teachers can then strategically introduce authoritative terms and resources at moments when they will extend and deepen rather than replace student ideas. Conversely, if no related terms are detected, the teacher can recognize that the concept has not yet been addressed and may require deliberate attention.

This kind of semantic curriculum coverage mapping directly supports the Knowledge Building principle of *constructive use of authoritative sources* (Scardamalia & Bereiter, 2006). Rather than treating authoritative terms as rigid checklists, the tool makes visible how student-driven inquiry is already converging toward canonical scientific ideas, while still valuing the diverse, informal language students bring to the discussion. Unlike earlier tools such as the Vocabulary Analyzer, which offered only a binary view (term present or absent), this approach provides a gradient that captures degrees of relatedness and partial understanding. It also reduces false negatives—cases where conceptual understanding is emerging but phrased differently—thus helping teachers recognize when students are ready to engage productively with authoritative knowledge sources in ways that advance collective inquiry.

Detection of related concepts beyond the curriculum

Knowledge Building emphasizes not just covering the official curriculum, but also going beyond—bringing in outside knowledge, making interdisciplinary connections, and pursuing unforeseen ideas (Scardamalia, 2002). Our semantic tool is particularly adept at detecting when students introduce *related concepts that were not in the provided glossary*. Because we analyze all content words, we inevitably encounter terms that are not on the teacher's original list. Some of these will be irrelevant or off-task, but others could be highly relevant to the topic and even extend it. The tool can cross-check the semantic similarity of these "extra" words to the entire set of curriculum terms to gauge their relevance. If a student-used word is semantically distant from all glossary terms, it might be an off-topic tangent (or something completely new). But if it is moderately close to one or more glossary terms, it could represent a valuable extension or enrichment of the topic. For example, in a unit on ecosystems, students might bring up "**climate change**" or "**pollution**" even if those terms were not in the original list of concepts for that unit. Our tool would identify these as prominent words in the discourse and might show they have some semantic proximity to terms like "ecosystem." The teacher's glossary did not include them because perhaps the curriculum did not explicitly address human environmental impact, but the students made that leap. The tool thus can flag such emergent topics for the teacher. This serves as evidence of *student-driven knowledge advancement*, connecting to KB principles of *idea diversity* and *knowledge building discourse* (Scardamalia & Bereiter, 2006) that crosses boundaries. It provides teachers and researchers a way to quantify and catalog how the community knowledge is extending beyond given parameters.

Illuminating cross-topic bridges in discourse

Another interesting analysis is how *cross-topic bridges* form. In many knowledge-building classrooms, students cycle through different investigation topics (e.g. one month on plants, another on ecology, another on human body). Ideally, ideas from earlier topics *permeate into later ones*, creating a continuous knowledge-building trajectory rather than isolated units. Our tool can detect when vocabulary from topic A reappears in topic B. Suppose a class studied *light* in physics and later *plant growth* in biology. If students make a connection (e.g., they mention "light" and "energy" in the plant discussion, relating to their physics knowledge), our semantic similarity measure might connect those uses to the earlier glossary terms from the light unit. We can produce an analysis that, for each pair of units or views, lists concepts that were carried over. For example: *Concept "energy" (from Unit 1 on Light) was referenced in Unit 2 on Plants (via words: light, sun, power)*. This shows cross-topic knowledge transfer. If one were to visualize the entire year's discourse as a network, each unit's glossary could be a cluster of nodes, and cross-topic bridges are edges between nodes of different clusters.

A healthy knowledge building environment might show many bridges (students leveraging prior knowledge to inform new inquiries), whereas a traditional unit-by-unit approach might show none. The tool thus supports *longitudinal tracking*, helping educators ensure continuity of learning. It also helps to document one of the promises of Knowledge Building pedagogy: that students will *"rise above"* and connect ideas across domains for deeper understanding.

For research, these analyses could be correlated with assessments of transfer or integrated understanding. This capability is somewhat analogous to what Authors (2019) described as *"crisscrossing topics"* – tracking student use of concepts beyond their grade-level expectations. Their analytic looked at how students incorporate

concepts across disciplinary boundaries and beyond curriculum guidelines, finding that KB students often *exceed* the bounds of the given content areas. Our tool automates the detection of such occurrences by spotlighting vocabulary that originates from outside the provided knowledge space but is relevant.

Collectively, these capabilities illustrate how a semantic analysis approach can yield a richer picture of discourse. It preserves much of what earlier analytics offered, but adds layers of meaning: recognizing paraphrases as evidence of understanding, surfacing student-introduced content, and detecting the evolution of the community's knowledge.

Classroom and research applications

The semantic analysis tool has broad applications in classroom practice and research. For teachers, it offers a new form of formative assessment, showing which key concepts have appeared in class discourse and which remain unexplored. By running analyses periodically, teachers can identify gaps and adapt instruction accordingly—supporting data-informed scaffolding consistent with Knowledge Building principles. The tool can also visualize community discourse to prompt collective reflection, helping students notice idea progress, missing concepts, and opportunities for deeper understanding.

For students, the tool provides formative feedback that fosters epistemic agency. Visualizations of concept networks make visible which ideas they have explored and where connections are lacking. This helps students take greater ownership of their learning, intentionally extend discussions, and contribute new perspectives. Rather than treating analytics as a checklist, students use them to guide inquiry and strengthen collaboration.

For researchers, the tool enables large-scale analysis of Knowledge Building discourse. Its semantic outputs quantify conceptual growth and integration, supporting studies on how scaffolds, teacher moves, or contexts influence idea development. It can also compare classes or correlate discourse patterns with learning outcomes, offering new ways to understand how semantic connectivity reflects and supports knowledge advancement.

Discussion and conclusion

In this paper, we introduced a novel semantic analysis tool tailored for Knowledge Building (KB) environments and demonstrated its potential to enrich both research and practice in collaborative learning. The semantic analysis tool advances key principles of Knowledge Building theory (Scardamalia & Bereiter, 1994, 2006) by making idea improvement visible through vocabulary and concept development. It operationalizes epistemic agency, allowing students to see how their contributions shape collective understanding and guiding teachers with ongoing semantic feedback that supports open-ended inquiry while ensuring core goals are met. The tool also embodies concurrent, embedded, and transformative assessment, turning evaluation into a collaborative process that informs further inquiry rather than judging it externally. By revealing idea diversity and highlighting connections among concepts, it promotes "Rise Above" synthesis, where communities integrate ideas into higher-order understanding. Ultimately, the tool bridges qualitative and quantitative approaches to discourse analysis, providing empirical evidence that students—when engaged in true knowledge building—can extend beyond the given curriculum, connect domains, and participate in progressive, expansive knowledge creation.

The successful integration of this tool would exemplify how learning analytics can be married with a strong pedagogical framework to produce actionable insights. Often, learning analytics suffer from focusing on what's easy to measure (clicks, counts) rather than what's educationally meaningful. Here, we take on the challenge of analyzing meaning – a complex thing to measure – because Knowledge Building demands it. We hope this inspires more "pedagogy-first" analytics development. Additionally, our approach demonstrates the value of combining *quantitative* and *qualitative methods*. The tool quantifies similarities, but its outputs are best interpreted with qualitative context (reading the notes, understanding the classroom narrative). It does not replace qualitative content analysis; rather, it guides where to look, making qualitative analysis more focused (akin to how digital humanities scholars use text mining to find themes in literature before close reading specific passages). Several limitations should be noted. The tool analyzes only textual data, so ideas expressed non-verbally (e.g., drawings or gestures) remain outside its scope, though future multimodal extensions could address this. Its performance also depends on the language model's quality—challenges may arise with highly specialized jargon, small datasets, or messy student text (e.g., spelling errors). While fine-tuning and preprocessing help, accuracy can vary. Nonetheless, these constraints are manageable through technical improvements and teacher awareness, and the tool's overall benefits outweigh its current limitations.

Despite these limitations, the work reported here is a step toward a *theory-aligned analytic innovation* that exemplifies the synergy between learning sciences and AI technology. The semantic analysis tool empowers educators and learners in Knowledge Building communities to better understand and guide their collective inquiry. It provides a new form of feedback that is content-aware, going beyond structural metrics to the level of meanings

and ideas. By integrating this tool into Knowledge Forum or similar platforms, we move towards a new generation of CSCL environments that not only connect learners but also *illuminate the knowledge they are building together*. We envision that as teachers and students use these semantic insights to reflect and adapt, classroom discourse will become more inclusive of important concepts, more connected across diverse ideas, and more visibly progressive toward deeper understanding. Such an outcome aligns deeply with Scardamalia and Bereiter's original vision for Knowledge Building – a community collectively improving its ideas with the help of supportive technology.

Future research will focus on evaluating, extending, and enhancing the tool. Empirical studies should assess its impact on classroom processes, concept coverage, and teacher practices through mixed-methods or quasi-experimental designs. Further, integrating automated feedback could help teachers respond more effectively to analytic insights by suggesting prompts or highlighting overlooked concepts, while maintaining teacher control and professional judgment. Finally, a student-facing dashboard could foster self-reflection by showing which concepts learners have (or have not) engaged with, encouraging more intentional contributions. Such developments would refine the tool's pedagogical value and broaden its applicability across disciplines. As we continue to refine the tool and study its impact, we anticipate new insights into how students learn together and how technology can respectfully and powerfully support that process. The journey of integrating advanced analytics with human-centric educational theory is just beginning, and the work presented here is a promising step on that path.

References

Bereiter, C., & Scardamalia, M. (2014). *Knowledge building and knowledge creation: One concept, two hills to climb*. In S. C. Tan, H. J. So, & J. Yeo (Eds.), *Knowledge Creation in Education* (pp. 35-52). Springer.

Burtis, P. J. (1998). *The Analytic Toolkit for Knowledge Forum*. [Software manual]. Institute for Knowledge Innovation and Technology, OISE/University of Toronto.

Chen, B., Scardamalia, M., & Bereiter, C. (2015). *Advancing knowledge-building discourse through judgment of promising ideas*. Paper presented at the Knowledge Building Summer Institute. (Example of Promising Ideas Tool use in Grade 3)

Chen, M., & Liu, X. (2016). *WordWhispers: Exploring automated suggestions for vocabulary enrichment in knowledge building*. Proceedings of the International Conference on Computers in Education. (Tool that recommends related words while students write).

Honnibal, M. (2017). spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing.

Hong, H.-Y., Scardamalia, M., Messina, R., & Teo, C. (2015). Fostering sustained idea improvement with principle-based knowledge-building analytic tools. *Computers & Education, 89*, 91–102. https://doi.org/10.1016/j.compedu.2015.08.012

Jung, Y., Zhu, X., Oshima, J., Oshima, R., Chen, B., Moon, J., ... & Jackson, T. J. (2025). Towards Actionable Collaborative Discourse Analysis: Bridging Advanced Computational Analysis with Practical Implementation. In *Proceedings of the 18th International Conference on Computer-Supported Collaborative Learning-CSCL 2025, pp. 472-480*. International Society of the Learning Sciences.

Khanlari, A., Zhu, G., & Scardamalia, M. (2019). Knowledge building analytics to explore crossing disciplinary and grade-level boundaries. Journal of Learning Analytics, 60(3), 60-75. https://doi.org/10.18608/jla.2019.63.9

Laferrière, T., Hamel, C., Perreault, M., & Allaire, S. (2016). *Lexical Analysis Tool for Knowledge Building Discourse*. In Proceedings of the International Conference on Collaborative Learning. (Compares student discourse to curriculum-based dictionaries) ()

Oshima, J., Oshima, R., & Matsuzawa, Y. (2012). Knowledge Building Discourse Explorer: A social network analysis application for knowledge building discourse. *Educational Technology Research and Development, 60*(5), 903–921.

Reimers, N., & Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*.

Resendes, M., Scardamalia, M., Bereiter, C., Chen, B., & Halewood, C. (2015). Group-level formative feedback and metadiscourse. *International Journal of Computer-Supported Collaborative Learning, 10*(4), 309–336.

Scardamalia, M. (2002). Collective cognitive responsibility for the advancement of knowledge. In B. Smith (Ed.), Liberal Education in a Knowledge Society (pp. 67-98). Open Court. (Introduces the 12 KB principles).

Scardamalia, M., & Bereiter, C. (1994). Computer support for knowledge-building communities. *The Journal of the Learning Sciences, 3*(3), 265–283. (Classic paper distinguishing KB from other approaches) ()

Scardamalia, M., & Bereiter, C. (2006). Knowledge building: Theory, pedagogy, and technology. In K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (pp. 97-118). Cambridge University Press.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*.

Wang, W., Wei, F., Dong, L., Bao, H., Yang, N., & Zhou, M. (2020). Minilm: Deep self-attention distillation for task-agnostic compression of pre-trained transformers. *Advances in neural information processing systems*, *33*, 5776-5788.

Zhang, J., Scardamalia, M., Lamon, M., Messina, R., & Reeve, R. (2007). Socio-cognitive dynamics of knowledge building in the work of 9-and 10-year-olds. *Educational Technology Research and Development*, *55*(2), 117–145. (Examines idea thread development over successive school years)

Zhu, G., & Kim, M. S. (2017). A review of assessment tools of Knowledge Building: Towards the norm of embedded and transformative assessment. Knowledge Building Summer Institute 2017.

 © ISLS

International Society of the Learning Sciences

An Item Response Theory Model for Multiple Resubmissions

Xiang Liu, Mo Zhang, Chen Li, Hongwen Guo
xliu003@ets.org, mzhang@ets.org, cli@ets.org, hguo@ets.org
Educational Testing Service

Amy J. Ko, Min Li
ajko@uw.edu, minli@uw.edu
University of Washington

Abstract: Item response theory models (IRT) are commonly used to model students' responses to test items. Under IRT, the associations among item responses are modeled by a latent variable. However, most IRT models are not suitable for modeling multiple resubmissions which are common in a learning environment. Students may be allowed to revise and resubmit multiple times after each submission. In this paper, we develop an IRT model for multiple resubmissions. The utility of this model is demonstrated through analyzing a real data set from a recently developed novel authentic assessment of programming skills.

Introduction

Item response theory (IRT; Baker & Kim, 2004; Hambleton & Swaminathan, 1985) is widely used to model students' responses to assessment items by relating observed item responses to an underlying latent proficiency. Under standard IRT assumptions, each student produces a single response to each item, and conditional on item parameters such as difficulty and discrimination, responses are locally independent and driven by a static latent trait. This framework has proven especially useful in summative assessment contexts, where responses are typically collected once per item and used to rank examinees along a unidimensional proficiency continuum.

In contrast, many contemporary learning environments—particularly computer-based formative assessments—allow students to revise and resubmit their responses multiple times, often with feedback provided after each submission. Such revise-and-resubmit workflows are increasingly common in online homework systems, programming assessments, and virtual learning environments, where the primary goal is to support learning rather than to measure end-state performance alone. In these settings, students' response processes generate rich multiple-attempt data that reflect not only their initial proficiency, but also how their performance evolves across attempts.

A growing body of psychometric research has recognized that traditional single-response IRT models are not well suited for such data. One prominent line of work extends IRT using sequential models, in which multiple attempts are treated as ordered responses within an item. For example, Bergner, Choi, and Castellano (2019) proposed sequential IRT models for constructed-response items with multiple attempts and demonstrated that explicitly modeling attempt sequences leads to better fit and ability estimation, particularly when students may stop attempting an item before reaching a correct response. Related work by Li et al. (2022) examined the performance of sequential 2PL IRT models under unstructured multiple-attempt data from virtual learning environments, showing that such models can recover ability reasonably well, but may exhibit bias when ability growth across attempts is substantial and not adequately captured.

Parallel developments have occurred in the context of multiple-choice items administered under answer-until-correct or multiple-attempt designs. For instance, Lu, Fowler, and Cheng (2025) developed a family of sequential IRT models that account for guessing and partial knowledge in multiple-choice, multiple-attempt settings, demonstrating that early incorrect attempts contain valuable psychometric information. Together, these studies provide strong evidence that multiple attempts are not psychometric noise, but rather an important source of information that should be modeled explicitly.

Despite these advances, most existing multiple-attempt IRT models do not directly parameterize students' propensity to improve across resubmissions. In sequential models, improvement is typically implicit in the attempt structure, while person parameters are still interpreted primarily as static proficiency. However, in authentic learning tasks—such as programming problems that require iterative debugging—students may differ not only in their initial proficiency, but also in how effectively they use feedback and revise their solutions. Treating these two aspects as indistinguishable may obscure meaningful individual differences in learning processes.

In this paper, we propose an extended IRT model for multiple resubmissions that explicitly separates these two components. In addition to a latent proficiency parameter, the model introduces a second person-level

 © ISLS

parameter that captures each student's propensity for improvement across successive attempts. The proposed model is designed for constructed-response tasks with revise-and-resubmit behavior and constrains improvement to be non-decreasing across attempts. We demonstrate the utility of the model using data from a computer-based Python programming assessment, showing how it provides a richer characterization of student performance than traditional IRT approaches.

A Python programming assessment

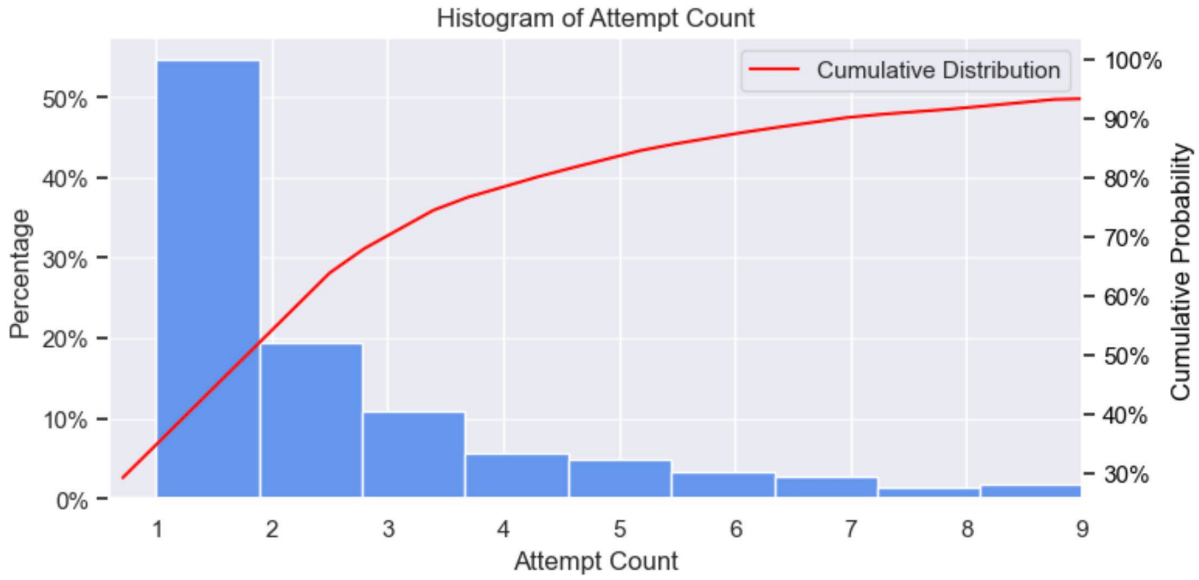
A test was designed to evaluate students' proficiency in basic Python programming. Participants were required to solve 14 to 15 programming tasks by submitting Python-coded solutions. These programs were executed against multiple sets of test cases. If a solution failed any test case, students received feedback indicating which cases failed and were allowed to revise and resubmit their code as many times as needed. The digital testing platform not only recorded the final responses but also logged detailed process data, including all keystrokes, program runs, window focus changes, and copy-paste actions.

The study involved 180 college students enrolled in an introductory Python course. All participants initially completed a set of 9 medium-difficulty items. Based on their performance, they were then directed to either an easier block of 6 items or a more challenging block of 6 items. Each solution was scored as correct (1 point) only if it passed all test cases; otherwise, it received 0 points. Many students opted to revise and resubmit their solutions until they achieved a correct result. The test had no time limit and allowed unlimited resubmissions. Nonetheless, most correct submissions occurred within the first five attempts (see Figure 1). As expected, the likelihood of a correct solution generally increased with more attempts. The data confirmed this trend, showing that students' total scores improved with additional revisions and resubmissions (see Figure 2).

There are some intricacies when modeling this type of item response. In contrast to the traditional single submission item response, multiple resubmissions add an extra dimension - the number of attempts. Consequently, it is no longer sufficient to just consider the propensity of getting items correct when measuring the students' performance, but we also need to take into account how many attempts a student makes before getting that correct response.

Figure 1

Total Number of Attempts Until a Correct Response



An extended IRT model for resubmissions

Let $X_{ij} = x_{ij}$ denote the item response from the ith student to the jth item, for $i = 1, 2, \ldots, N$ students and $j = 1,2, \ldots, K$ items. If the response is correct, $x_{ij} = 1$; otherwise, $x_{ij} = 0$. Under a two-parameter logistic IRT model, the probability of a correct response is given by

 © ISLS

$$P(X_{ij} = 1|\theta_i) = \frac{\exp\left(a_j(\theta_i - b_j)\right)}{1 + \exp\left(a_j(\theta_i - b_j)\right)}, \quad (1)$$

where θ_i describes the proficiency level of the *i*th student, and a_j and b_j are the item discrimination and the difficulty parameters for the *j*th item. Sometimes, especially under small sample sizes, the item discrimination parameters a_j are constrained to be the same for all $j = 1,2,\dots,K$, i.e. $a_j = a, \forall j \in \{1,2,\dots,K\}$. Then the IRT model becomes

$$P(X_{ij} = 1|\theta_i) = \frac{\exp\left(a(\theta_i - b_j)\right)}{1 + \exp\left(a(\theta_i - b_j)\right)}, \quad (2)$$

To accommodate item responses with multiple resubmissions, we extend the IRT model. Let $X_{ijt} = x_{ijt}$ denote the *t*th submission from the *i*th student to the *j*th item. $x_{ijt} = 1$ if the submission is correct and $x_{ijt} = 0$ if it is incorrect. In addition to the latent proficiency parameter θ_i, We introduce another person parameter γ_i which describes the propensity of the *i*th student getting improved probability of a correct response during multiple resubmissions. Furthermore, we constrain the parameter to be nonnegative, i.e. $\gamma_i \geq 0$, so that a student's correct response probability is non-decreasing in multiple resubmissions. Under the model, the probability of a correct response is now given by

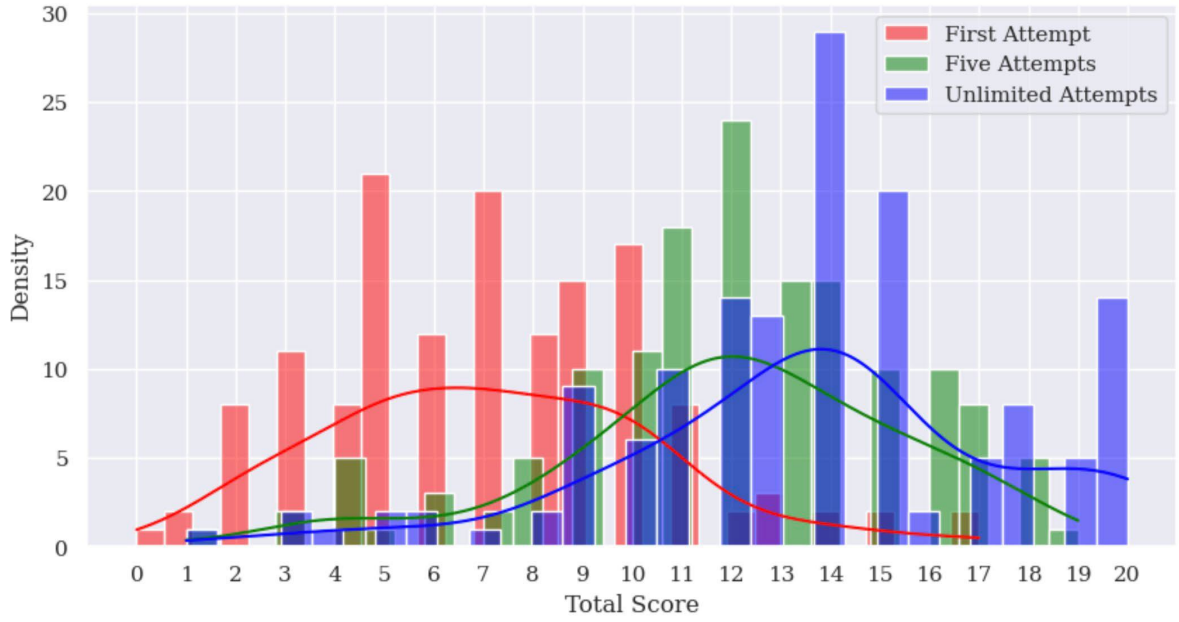
$$P(X_{ijt} = 1|\theta_i, \gamma_i) = \frac{\exp\left(a(\theta_i - b_j + (t-1)\gamma_i)\right)}{1 + \exp\left(a(\theta_i - b_j + (t-1)\gamma_i)\right)}, \quad (3)$$

Where the person proficiency parameter, θ_i, is assumed to follow a standard normal distribution $\theta_i \sim N(0,1)$.

Under this extended IRT model, a student is described by a pair of latent variables – (θ_i, γ_i). For two students who may be located differently in this two-dimensional space, if $\theta_i \geq \theta_j$, the *i*th student is more proficient than the *j*th student. As a result, the more proficient student is more likely to get items correct on the first submission. But the *j*th student may be more likely to improve during subsequent resubmissions, if $\gamma_i < \gamma_j$. Notice that the improvement is in the logit scale, not the probability scale.

Figure 2

Total Score by Number of Attempts



 © ISLS

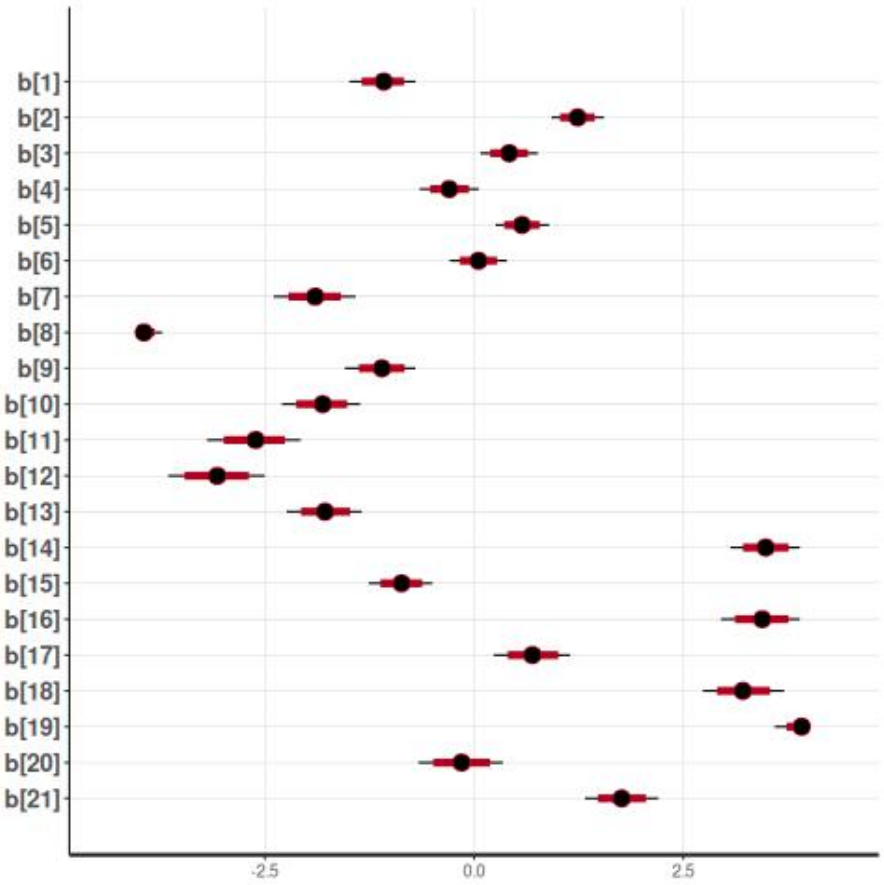
Real data analysis

To demonstrate the utilities of our proposed approach, we fit the model to the Python programming assessment dataset. The model parameters are estimated using a Bayesian approach. The Bayesian estimation is carried out using STAN (Gelman et al., 2015) - a general-purpose Bayesian inference software. Interested readers may find the code in the appendix.

We used 5000 iterations in the Markov chain Monte Carlo (MCMC; e.g. Liu and Johnson, 2019) algorithm, of which half were treated as burn-ins and the rest were used for summarizing posteriors. Figure 3 shows the estimated item difficulty parameters. The difficulty levels of the items are clearly differentiated. The item difficulties of the easier items are estimated to be around –2.5, those of medium level items are around 0, and the more difficult items have item difficulties above 2.5.

Figure 3

Posterior Credible Intervals of Item Difficulty Parameters

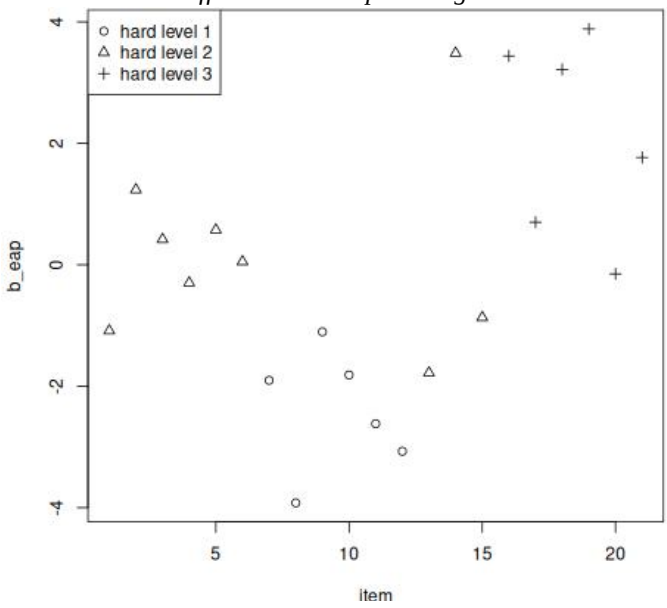


In Figure 4, the estimated item difficulty parameters are compared against the expert judged difficulty levels during the item development. The results of fitting the data to the extended IRT model empirically validated the expert judgement.

 © ISLS

Figure 4

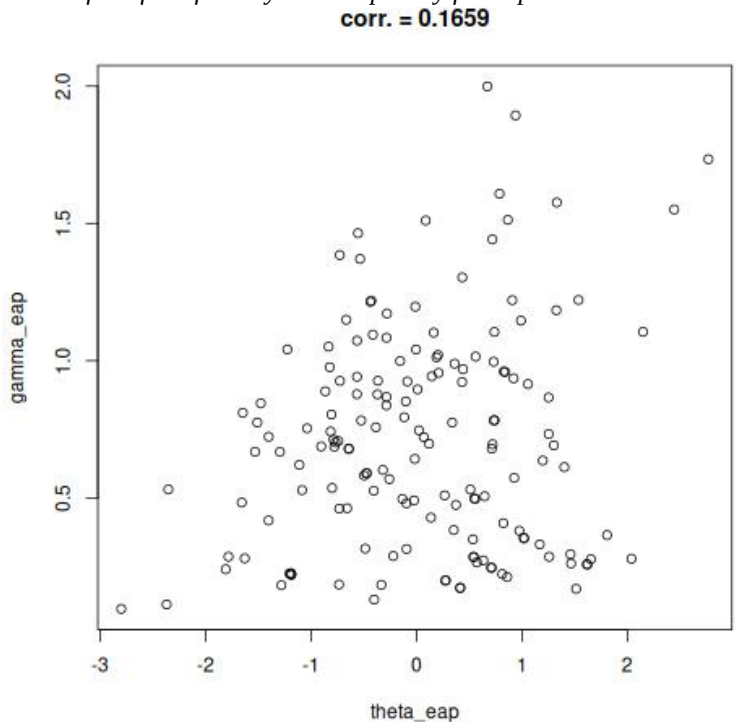
Estimated Item Difficulties Vs. Expert Judgement



In a traditional IRT model, the main utility is to rank order students along a single dimensional latent continuum. However, the extended IRT model offers some important additional utilities, such as profiling students on a two-dimensional space which may help gauge students' proficiency levels together with their tendencies to improve over multiple resubmissions. Figure 5 shows that, even though the overall correlation between programming proficiency and the propensity of improvement is not particularly high at 0.1659, there are some recognizable patterns for the extremes. The lower proficient groups tend to have lower propensity for improvement during resubmissions, while the most proficient groups tend to have higher propensity for improvement.

Figure 5

2D Profile of Proficiency and Propensity for Improvement



Discussion

In this paper, we proposed an extended IRT model for modeling multiple resubmissions. The model is demonstrated in the context of analyzing a computer programming assessment dataset. However, the introduced model can be used in other types of classroom assessments where multiple resubmissions are common. The results from fitting the model can help practitioners refine items and provide opportunities for delivering individualized interventions.

While the proposed model offers a useful framework for capturing students' improvement across multiple resubmissions, it simplifies the underlying dynamics by assuming a linear improvement propensity. Students' revise-and-resubmit behaviors can be far more complex. For instance, some students may exhibit rapid improvement early on and plateau later, while others may show sporadic or delayed gains. Factors such as motivation, feedback quality, cognitive load, and prior knowledge can all influence the trajectory of improvement in non-linear ways.

The current model assumes that the probability of a correct response increases linearly with each additional attempt, governed by a single parameter γ_i. This assumption, while computationally convenient, may be too restrictive for capturing the nuanced patterns observed in authentic learning environments. Future research could explore more flexible formulations, such as non-linear growth functions, piecewise models, or hierarchical structures that allow for varying improvement rates across different items or student subgroups.

By incorporating richer behavioral dynamics, extended models could offer deeper insights into learning processes and better support adaptive instructional strategies. Such developments would enhance the interpretability and applicability of IRT-based approaches in formative assessment contexts.

References

Baker, F.B., & Kim, S.-H. (2004). Item response theory: Parameter estimation techniques. CRC Press.

Bergner, Y., Choi, I., & Castellano, K. E. (2019). Item response models for multiple attempts with incomplete data. Journal of Educational Measurement, 56(2), 415–436. https://doi.org/10.1111/jedm.12214

Gelman, A., Lee, D., & Guo, J. (2015). Stan: A Probabilistic programming language for Bayesian inference and optimization. Journal of Educational and Behavioral Statistics, 40(5), 530-543.

Hambleton, R.K., & Swaminathan, H. (1985). Item response theory: Principles and applications. Kluwer-Nijhoff Publishing.

Li, Z., Huggins-Manley, A. C., Leite, W. L., Miller, M. D., & Wright, E. A. (2022). Assessing ability recovery of the sequential IRT model with unstructured multiple-attempt data. Educational and Psychological Measurement, 82(6), 1203–1224. https://doi.org/10.1177/00131644211058386

Lu, Y., Fowler, J., & Cheng, Y. (2025). A family of sequential item response models for multiple-choice, multiple-attempt test items. Psychometrika. Advance online publication. https://doi.org/10.1017/psy.2024.18

Liu, X., & Johnson, M.S. (2019). Estimating CDMs using MCMC. In M. von Davier & Y.-S. Lee (Eds.), Handbook of Diagnostic Classification Models (pp. 629-646). Springer.

 © ISLS

Appendix

Code for fitting the extended IRT model in STAN

```
data {
        int<lower=1> J; // # students
        int<lower=1> K; // # items
        int<lower=1> N; // # obs
        int<lower=1> M; // # max attempts
        int<lower=1, upper=J> jj[N]; // student for obs.n
        int<lower=1, upper=K> kk[N]; // item for obs.n
        int<lower=1, upper=M> mm[N]; // attempt for obs.n
        int<lower=0, upper=1> y[N]; // obs
}
parameters {
        real theta[J];
        real<lower=-4, upper=4> b[K];
        real<lower=0, upper=4> gamma[J];
        real<lower=0.1, upper=3> a;
}
model {
        theta ~ std_normal();
        for(n in 1:N)
                y[n] ~ bernoulli_logit(a*(theta[jj[n]] - b[kk[n]] + (mm[n] - 1)*
        gamma[jj[n]]));
}
```

 © ISLS

International Society of the Learning Sciences

Estimating Teacher Orchestration Load in Co-located and Online CSCL: Multimodal Data Analysis

Lubna Hakami, Ishari Amarasinghe, Davinia Hernandez-Leo
lubna.hakami01@estudiant.upf.edu, ishari.amarasinghe@upf.edu, davinia.hernandez-leo@upf.edu,
Universitat Pompeu Fabra

Abstract: Teacher orchestration load has been an area of interest in the field of Computer-Supported Collaborative Learning (CSCL) as it reflects the real-time cognitive and emotional burden teachers experience. Orchestrating CSCL activities, especially in different learning environments, may impose stress on teachers. This study investigates teacher orchestration load when using a scripting tool, focusing on two learning settings: co-located and online. Multimodal data combined self-reports with physiological measures to capture teacher orchestration load. In this context, electrodermal activity (EDA) was used to capture physiological indicators of teacher orchestration load. EDA peaks corresponded with self-reported stressful moments, serving as an additional indicator for estimating orchestration load. Results revealed higher orchestration load in co-located settings compared to online. Moreover, factors such as the number of students and task type affect teacher orchestration load in both settings. Understanding these factors can support the design of teacher-supporting technologies that may assist in balancing teacher workload.

Introduction

In the domain of Technology-Enhanced Learning (TEL) and CSCL, orchestration is the real-time management of classroom activities using various technological tools (Dillenbourg et al., 2011; Giano & Patton, 2023). This involves the real-time adjustment and coordination of the technological tools, resources, tasks, and social design used to effectively address the evolving needs of students and adapt to the dynamics of the learning environment. This study focuses on orchestration as the run-time coordination of CSCL activities by teachers. Moreover, real-time orchestration may impose a considerable burden on teachers, including cognitive, decision-making, technological, emotional, and physical dimensions. Understanding the orchestration is crucial as it could impact the effectiveness of the learning environment. This brings us to the concept of teacher orchestration load, which is “the effort necessary for the teacher – and other actors – to conduct learning activities" (Cuendet et al., 2013).

Despite its importance, estimating orchestration load remains challenging. Earlier studies often treated it as a “black box” (Amarasinghe et al., 2021a) due to the lack of valid measurement tools. Some used intrusive methods such as eye-tracking glasses (Prieto et al., 2017, 2018), which can interfere with natural teaching behaviors, while others focused mainly on co-located settings. More recent work has adopted multimodal data approaches, integrating physiological, behavioral, and self-reported data, to gain richer insights into orchestration processes (Prieto et al., 2015a; Crespi et al., 2022; Hakami et al., 2024).

Building on this line of research, the present study uses non-invasive wearable sensors to collect Electrodermal Activity (EDA) data as a physiological indicator of teacher stress and workload. EDA, which measures changes in skin conductance linked to emotional arousal, has been widely used in health, sports, and education research but rarely applied to teachers. To contextualize physiological signals, the study triangulates EDA peaks with qualitative data, teachers’ self-reported stressful moments, observation notes, and dashboard screen recordings, to interpret what triggers orchestration load during CSCL activities.

Therefore, the study aims to address a gap in the literature by extending the investigation of Orchestration Load in online learning environments and demonstrating how multimodal analytics can reveal underlying factors contributing to teacher stress and workload. This examination becomes more potent with the growing use of multimodal data analytics in educational technology research (Blikstein & Worsley, 2016; Sharma & Giannakos, 2020). The main research question the study aims to address is: *What can multimodal data reveal about teacher orchestration load in co-located and online CSCL settings?* The remainder of this paper is structured as follows. Section 2outline methods, including data collection and analyses. Section 3 presents results, Session 4 discusses the study findings, and Section 5 concludes the study with limitations and directions for future research.

Methods

This study employed a *mixed-method approach***,** following a *single-subject research design* (SSRD) to examine *teacher orchestration load* in scripted CSCL activities across *co-located and online settings*. By combining *quantitative physiological data* with *qualitative teacher perceptions*, it provided a comprehensive view of

 © ISLS

orchestration load as both an objective and subjective phenomenon. The SSRD was chosen because orchestration load varies by individual and context, allowing detailed analysis of a single teacher's experience (a female teacher from a Spanish public university) across *20 CSCL* **sessions** without needing a separate control group. This design enabled in-depth exploration of how orchestration load changes across settings and what factors influence it.

Procedures and materials

We collected multimodal data while the teacher was monitoring the CSCL activities in both co-located and online settings (N = 20). Objective data collected included electrodermal activity (EDA) from a Shimmer3 GSR+ sensor, also known as Galvanic Skin Response (GSR). This study focuses on EDA peaks as a physiological indicator of increasing the orchestration load in CSCL activities. The sensor was strapped (see Figure 1) to the teacher's non-dominant hand, specifically, the left forearm, and electrodes were placed on the inner side of the left wrist (Dooren et al., 2012). The sensor monitored the teacher non-invasively. Moreover, log data extracted from the orchestration tool's database was collected to extract more details about the activities. Dashboard screen recordings were collected when the teacher orchestrated the CSCL activities to analyze the teacher's actions using the dashboard. Observation notes and responses to post-activity questionnaires were also collected to collect details on stressful moments. The Ethical approval for this study was obtained from a university ethics review board.

Figure 1
Shimmer3 GSR+ (Galvanic Skin Response)



After removing sessions with missing data, 12 CSCL sessions (six for co-located and six for online settings) were used for further analysis (see Table 1). The orchestration tool used in this study is a web-based CSCL tool that follows the pyramid CLFP based on scripted collaborative learning activities (name removed for blind review). It enables teachers to create CSCL activities in diverse learning settings (co-located, online, and hybrid). During a CSCL activity, students individually respond to the task provided by the teacher. Then, they are automatically assigned to groups to rate the other answers from students. Lastly, they collaborate with their group members and agree to write appropriate answers. The tool has a teacher-facing dashboard (see Figure 2) designed for teachers to control and monitor the activity flow in real-time and take pedagogical actions. The dashboard also shows real-time students' activity participation and contains controls that enable activity adjustment (i.e., increase time, pause activity, skip phase, and end activity when needed). The dashboard has a flagging feature that allows teachers to flag (mark) student answers and present them in the dashboard interface for discussion (Dieckmann et al., 2022). In addition, the dashboard has three tabs (see Figure 3): Student Task Tab, which displays the task provided to the students, Response Related Tab, which presents the individual answer from each student, the highly voted answer, and the final answer of a group, depending on the phase of the activity, and Participation Related Tab, which presents the participation differences between groups. The dashboard visualizes students' CSCL activity participation in real-time and is aimed at supporting teachers' real-time regulation of the activity by showing relevant information (e.g., group participation and collaboration or task progress) and alerting teachers (i.e., using specific alerts triggered by learning analytics indicators) on relevant interventions.

Figure 2
Teacher-Facing Dashboard for The Orchestration Tool

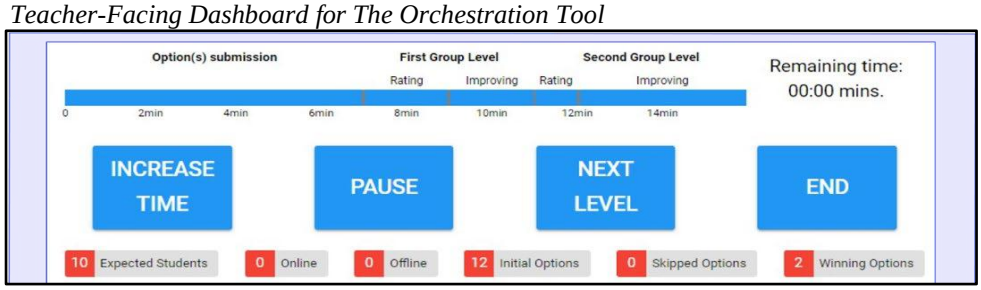


Figure 3

The Tabs in The Dashboard:(A)Student Task Tab, (B)Response Related Tab, (C)Participation Related Tab.

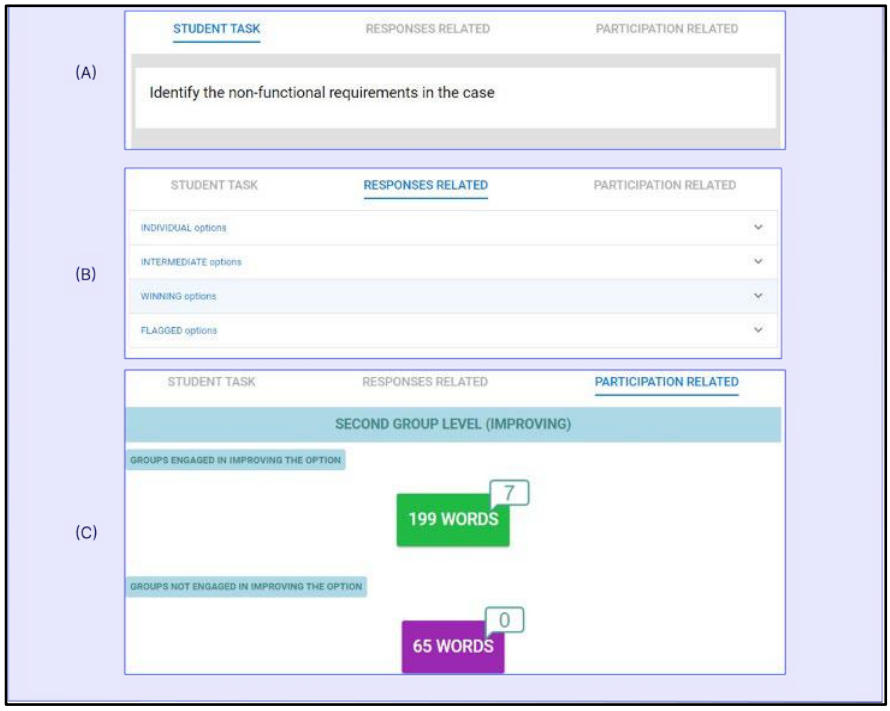


Table 1

The Pyramid Activities in Co-Located and Online Sessions, Along With the Activity Duration, the Number of Students Participating in the Activity, and the Task Given to Students.

Session No.	No.student*	Duration (min)	Task	Task Type
Co-located				
C.1	59	19	What is considered plagiarism and why is it important to know it in university?	Comprehension
C.2	77	19	What is considered plagiarism and why is it important to know it in university?	Comprehension
C.3	62	19	What is considered plagiarism and why is it important to know it in university?	Comprehension
C.4	10	19	Identify the non-functional requirements in the case	Analysis
C.5	14	15	Indicate a maximum of three positive and three negative aspects of technological evolution.	Evaluation
C.6	73	12	What is considered plagiarism and why is it important to know it in university?	Comprehension
Online				
O.1	15	9	Which factors should be considered when considering the implementation of learning analytics?	Synthesis
O.2	16	7	Analyzing a scenario to identify requirements	Analysis
O.3	15	9	Analyzing a scenario to identify requirements	Analysis
O.4	15	8	Analyzing errors in code	Analysis
O.5	14	17	Based on your experience, what "computer" features are relevant to "support" cooperation or collaboration?	Knowledge
O.6	16	18	List differences between a LMS an a MOOC platform	Comprehension

The students were different in co-located and online settings, but also across some sessions within each setting, thereby reducing the potential for bias in the teacher's attention that could arise from familiarity with specific students.

Data analysis

 © ISLS

ConsensysPRO software was used to extract EDA data from the sensor. Then, a tool called EDA Explorer was used (Taylor et al., 2015) to automatically identify EDA peaks. The EDA peaks reflect the increased emotional arousal experienced under stress (Salim et al., 2020). Focusing on EDA peaks as a potential indicator of increasing the orchestration load, as it provides an objective measure of the teacher's physiological response to the demands of managing CSCL activities. Therefore, this study correlates each EDA peak with the stressful moments reported by the teacher.

Consequently, the stressful moments (Appendix A) are analyzed through a qualitative content analysis approach. These stressful moments were grouped into three categories, namely, Technological difficulties (e.g., Dashboard Problems), Actions by students (e.g., Noises from the students or Answers), and Time-related issues (Hakami et al., 2023). Through the analysis of the data, an additional category emerged: Teacher-related, and this category is related to teachers' actions, such as issues in the activity design. The other data (i.e., dashboard screen recordings and observation notes) provided additional details of the context, helping us to identify information that was not captured in self-reported data alone. Therefore, these stressful moments were analyzed in detail using the dashboard screen recordings and the observation notes, linking them with EDA data.

Correlation analysis was performed to investigate factors affecting teacher orchestration load. The variables used in the analysis were relevant to the teacher orchestration load in both co-located and online settings. Specifically, a dependent variable extracted from the sensor (EDA peaks), and three independent variables extracted from log data (number of students, activity duration, and task type used in the activity). The tasks used in CSCL activities are categorized based on Bloom's taxonomy (i.e., knowledge, comprehension, analysis, synthesis, and evaluation) (Bloom, 1956). All variables were chosen based on a previous study (Hakami et al., 2024), which identified these variables as factors that affect teacher orchestration load.

Results

In the following, we present the main results obtained in this study to answer the research question: *What can multimodal data reveal about teacher orchestration load in co-located and online CSCL settings?* To address the research question, we conducted two different analyses. First, we identified the stressful moments and triangulated them with EDA peaks. Second, we conducted the correlation analysis between EDA and contextual factors that emerged from the learning design (activity's duration, and task type) and context, such as the number of students who participated in sessions, to investigate their influence on orchestration load in CSCL. The results of these analyses are presented in the following sections.

Stressful moments and EDA analysis

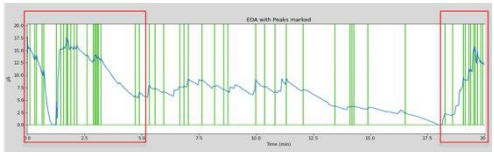
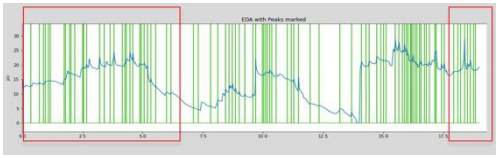
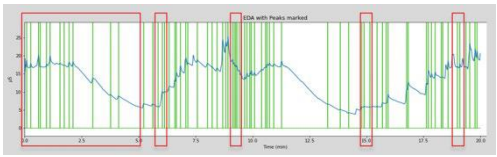
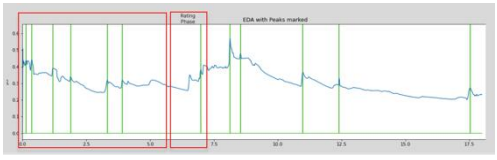
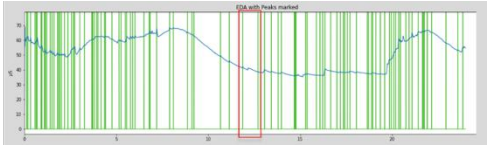
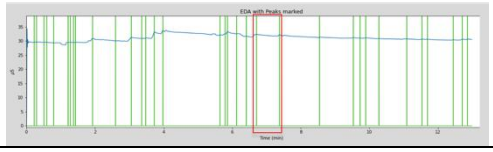
In Tables 2 and 3, we present the results related to teacher-reported stressful moments from the self-reported questionnaire, along with EDA signals highlighting the related stressful moments. First, considering the responses received to the self-reported questionnaire, it was found that there were nine stressful moments in the co-located setting and five stressful moments in the online setting. In the co-located setting, teachers reported three stressful moments related to student actions, such as managing classroom disruptions. Two stressful moments related to the technological issues and difficulties using the orchestration tool on a tablet. In contrast, these issues were absent in the online setting; instead, there were two moments related to teacher-related issues, such as configuration problems (forgetting to publish the activity). Lastly, in both settings, the teacher experienced stressful moments related to managing activity time, such as increasing or reducing the time for activities, specifically four moments in the co-located settings and three in the online settings.

We detected a total of 402 EDA peaks in the co-located setting and 46 EDA peaks in the online setting. The higher number of peaks in the co-located setting suggests that the teacher experiences more frequent and intense stress in co-located classrooms when compared to online settings. In this study, EDA peaks are often correlated with self-reported stressful moments, such as managing student answers, handling time constraints, or dealing with classroom noise. Also, the difficulty in pressing some buttons in the dashboard was corroborated by both the EDA data and screen recordings in sessions 5 and 6 (co-located setting). In the Online setting, although EDA peaks were less pronounced (i.e., in sessions 4 and 6), the other sessions still aligned with reported stressful moments. Further details on the results presented above are reported in Tables 2 and 3.

Table 2

Multimodal Analysis of Teachers' Stressful Moments in Co-located CSCL Sessions

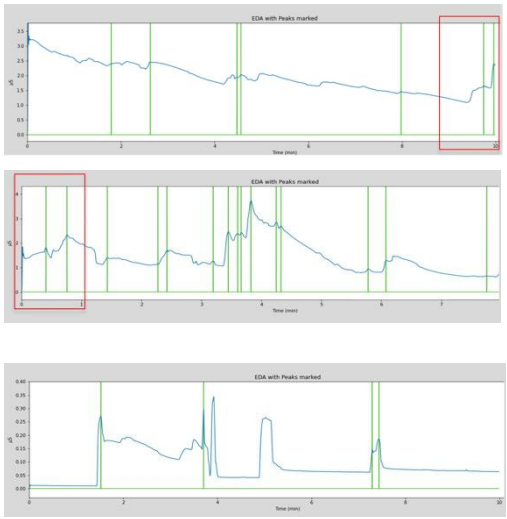
Explanation reported by the teacher for the stressful moment(s)	Figures
Session#1 Stressful Moment #1.1:	

Explanation reported by the teacher for the stressful moment(s)	Figures
The teacher reported a stressful moment during the activity due to the high number of answers, resulting in more effort in choosing which answer should be presented/debriefed at the end of the activity. When the screen recordings were analyzed, it was found that the teacher was looking at the answers in the first five minutes (left red box) and the last two minutes (right red box) of the activity. As seen in the figure, these stresfull moments align with increased EDA peaks. (Related Category: Actions by students)	
Session #2:Stressful Moment #2.1: The teacher reported a stressful moment related to the students' answers. After analyzing and reviewing the dashboard screen recordings, we found the teacher read students' answers and used the flag feature in the dashboard during the first six minutes (left red box) and the last minute (right red box) of the activity. (Related Category: Actions by students).	
Stressful Moment #2.2: The teacher reported another stressful moment about extending or not extending activity time. However, identifying the exact moment is challenging, as it was not captured clearly in the screen recording or observation notes. It could be at any time of the activity, as we can see the EDA peaks over the activity. (Related Category: Time management).	
Session #3:Stressful Moment #3.1: The teacher mentioned that during the first phase of the activity (the first five minutes), she noticed students needed more time when the individual submission time was finished. However, she managed to increase the time. (Related Category: Time management).	
Stressful Moment #3.2: The teacher reported that after each phase finished (the phases finished at 6, 9, 15, and 18 minutes), the students started to talk, and the class became noisy (Related Category: Actions by students).	
Session #4: Stressful Moment #4.1: First, when the teacher did not realize she needed to extend more time for the first phase (i.e., the first six minutes). (Related Category: Time management).	
Stressful Moment #4.2: Second, during the rating phase, the time was short, and many students could not complete the rating. The teacher mentioned that she felt bad that she could not extend the time. (Related Category: Time management).	
Session #5: Stressful Moment #5.1: The teacher reported that it was difficult to press the “Next Level” button on the dashboard. (Related Category: Technological difficulties).	
Session #6: **Stressful Moment #6.1:** Similarly, the teacher reported that it was difficult to press the “Next Level” button on the dashboard. However, she said this moment was not really stressful.	

Explanation reported by the teacher for the stressful moment(s)	Figures
(Related Category: Technological difficulties).	Note: The red boxes in the figures highlight the time intervals where there are more EDA peaks.

Table 3

Multimodal Analysis of Teachers' Stressful Moments in Online CSCL Sessions.

Explanation reported by the teacher for the stressful moment(s)	Figures
Session #1: **Stressful Moment #1.1:** The teacher mentioned one stressful moment, but it was not highly stressed. She was running out of class time and needed to reduce the activity time, so she pressed the "End" button on the dashboard to finish the activity. (Related Category: Time management). **Session #2:** **Stressful Moment #2.1:** The teacher reported that at the beginning of the activity, she noticed that the task shared with students was not the planned one. However, after reading the task quickly, she found that the task was correct. (Related Category: Teacher-related). **Session #3:** The teacher did not report any stressful moments. **Session #4:** **Stressful Moment #4.1:** The teacher mentioned that he had a stressful moment (not highly stressful) because she did not notice the activity time finishing. However, there were no peaks at the end of the activity. (Related Category: Time management). **Session #5:** **Stressful Moment #5.1:** The teacher reported that at the beginning, she forgot to publish the activity (publishing the activity was needed to start the activity in the tool). (Related Category: Teacher-related). **Session #6:** **Stressful Moment #6.1:** The teacher reported two stressful moments. First, although the teacher reported that she was not paying attention to the time, there were no EDA peaks. Second, the improving phase finished before the expected time. (Related Category: Time management).	 Note: The red boxes in the figures highlight the time intervals where there are more EDA peaks.

Factors influencing teacher orchestration load

The study aimed to investigate how teacher orchestration loads differ in each learning setting (co-located and online). However, due to the differences in the number of students (295 vs. 91), session duration (103 vs. 68 minutes), and the types of tasks used, we further investigate whether these factors may influence teacher orchestration load in each setting. Moreover, a previous study found that these factors (number of students, duration of the activity, and task type) affect teacher orchestration load in a co-located setting (Hakami et al., 2024). The current study investigates whether these factors influence the online setting. Also, compare their impact between co-located and online settings; thereby, correlation analysis (Table 4) was used to study the effect of the three factors.

In the co-located setting, three factors affect teacher orchestration load. The session duration of the activity showed a strong positive significant correlation (r = 0.89*, p = 0.01), indicating that longer activities considerably increase the teacher orchestration load. The number of students had a strong positive correlation (r = 0.74, p = 0.09), suggesting that more students may contribute to a higher orchestration load. Task type, had a weak, non-significant correlation (r = 0.21, p = 0.68), In the online settingsthe number of students had a strong positive and statistically significant correlation (r = 0.85*, p = 0.03), indicating that the continuous monitoring and managing of more students online substantially increased the teacher orchestration load; therefore, it became more demanding as the number of students increased. Task type showed a weak, non-significant correlation (r = 0.04, p = 0.93). The duration of the activity showed a weak negative correlation (r = -0.07, p = 0.89), implying that the length of the activity did not impact the orchestration load in the online setting. The teacher tends to focus on different aspects of the activity, such as monitoring student collaboration or focusing on the epistemic aspect (related to the learning task), rather than adjusting the time.

Table 4

Correlations for Co-Located and Online Settings

Dependent Variable	Independent Variable	Co-located		Online	
		Pearson's r	p-value	Pearson's r	p-value
No.peaks	Duration (min)	**0.89***	0.01	-0.07	0.89
	No.students	0.74	0.09	**0.85***	0.03
	Task Type	0.21	0.68	0.04	0.93

Discussion

This study advances the field of CSCL by showing how multimodal data can provide a richer and a comprehensive understanding of teacher orchestration load.Unlike previous studies that relied on a single modality (Prieto et al., 2015b), this study used non-invasive sensors to collect the EDA data in real-time without disrupting classroom orchestration. By linking EDA data with dashboard screen recording, observation notes, and self-reports, the study identified stress triggers and contextualized moments of high orchestration load. The findings show that multimodal analytics is a valid and comprehensive approach to assessing teacher orchestration load in scripted CSCL contexts. The comparative analysis between co-located and online settings revealed distinct sources of stress:

In co-located settings, teachers frequently reported stress related to Actions by students and Technological difficulties categories that were not in the online setting. The physical presence in a classroom needs immediate interventions for disruptions, managing classroom noise, or interacting with students, adding to the load on teachers, especially when controlling CSCL activities. Additionally, the usability of the orchestration tool's dashboard on a tablet in the co-located setting posed challenges. The teacher reported difficulty using the dashboard, which may be due to the tablet's touch sensitivity or screen size, leading to an added load on the teacher during activity management. A recent study also reported that technical issues affect teacher orchestration load in collaborative hybrid settings (Ortega-Arranz et al., 2024). In contrast, in the online setting, the stressful moments were related to the Teacher-related category, mainly due to configuration issues, such as forgetting to publish the activity, which prevented students from starting the task and may cause stress for the teacher. The Time-related issues were a common source of load in both settings.

Moreover, the study found that teacher orchestration load was higher in co-located settings. The division of teacher attention between physical and digital spaces might contribute to an increase in teacher stress in the co-located setting (Hakami et al., 2023). Understanding the specific triggers of orchestration load can inform the design of more supportive CSCL tools and dashboards.

In addition, the study also identified three contextual factors, i.e., number of students, activity duration, and task type, that influence orchestration load beyond the learning setting (co-located vs. online). Specifically, the *number of students* showed a positive relationship with *EDA peaks*, indicating that larger class sizes increase orchestration load. This supports prior studies showing that more students heighten stress (Huang et al., 2022) and classroom management challenges (Biggs, 1999). Regarding task type, our results indicated that task type was found to be a factor affecting the load in both co-located and online settings. Previous research by Amarasinghe et al. (2021b) has shown that certain task types may demand more targeted teacher interventions to achieve learning goals, thereby increasing workload. Finally, there was a difference in the duration factor, which was found to be a significant factor affecting teacher load in co-located settings when compared to online settings. This difference supports that in co-located classrooms, longer activity durations often involve direct interaction with students, managing real-time classroom dynamics, moving around the classroom, or addressing immediate

student needs. Thus, the activity's duration could impact the orchestration load in co-located settings (Hakami et al., 2024).

Conclusions and future work

This study built upon the previous studies from (Amarasinghe et al., 2021c; Crespi et al., 2022), which proposed triangulating EDA with self-reported data to assess the teacher orchestration load in CSCL activities. The finding shows that the triangulation of EDA with self-reported data can be a valid approach for estimating teacher orchestration load in both co-located and online settings. This was achieved by analyzing the stressful moments reported by the teacher in the questionnaire and correlating them with EDA peaks extracted from the sensor. Overall, the EDA peaks are often associated with the stressful moments in which the teacher reported increased stress or the load in the CSCL activities. Thus, the findings indicate that EDA was a promising physiological indicator to estimate teacher orchestration load in CSCL activities.

Furthermore, the findings provide insight into the differences in teacher orchestration load between the settings. The study highlights the greater orchestration load experienced by teachers in co-located CSCL settings. This study highlights the distinct challenges and stress faced by teachers in various learning environments, providing insights for designing more effective and supportive CSCL tools and dashboards. In addition, we have also identified some factors (i.e., the number of students, duration of the activity, and task type) that influence teacher orchestration load.

While this study is focused on a single subject, some of the teacher workload observed aligns with previous research. For example, the number of students may impact the teachers' physiological stress (Huang et al., 2022) and increase classroom management challenges (Biggs, 1999). In addition, the technical issues affect teacher orchestration load (Ortega-Arranz et al., 2024). This shows that certain dependencies in orchestration load, such as managing real-time student interactions, handling technological difficulties, teacher-related issues, or managing time, may be consistent in different teaching contexts. The current study provides an example of how, in this specific case, orchestration load dependencies emerged based on student number, task type, and activity duration. As more cases are studied, these insights can inform the development of orchestration tools that support the needs of different learning environments.

The study acknowledges several limitations. It focuses on a single female teacher, which limits generalizability and overlooks possible gender, experience, and familiarity effects on orchestration load. The study also examines only one collaborative learning flow pattern (pyramid CLFP); therefore, findings may not apply to other collaborative designs. Reliance on self-reported questionnaires introduces potential bias; future research should include qualitative methods like interviews or focus groups for deeper insights. Additionally, future work should explore hybrid learning contexts that combine online and face-to-face teaching, as these add complexity to orchestration. Finally, understanding how variables such as class size, activity duration, and task type affect teacher stress can guide interventions to reduce orchestration load.

References

Amarasinghe, I., Hernández-Leo, D., & Ulrich Hoppe, H. (2021a). Deconstructing orchestration load: comparing teacher support through mirroring and guiding. *International Journal of Computer-Supported Collaborative Learning*, 16(3), 307-338. https://doi.org/10.1007/s11412-021-09351-9

Amarasinghe, I., Hernández-Leo, D., Theophilou, E., Roberto Sánchez Reina, J., Quintero, R.A.L. (2021b). Learning Gains in Pyramid Computer-Supported Collaboration Scripts: Factors and Implications for Design. In: Hernández-Leo, D., Hishiyama, R., Zurita, G., Weyers, B., Nolte, A., Ogata, H. (eds) Collaboration Technologies and Social Computing. CollabTech 2021. Lecture Notes in Computer Science(), vol 12856. Springer, Cham. https://doi.org/10.1007/978-3-030-85071-5_3

Amarasinghe, I., Vujovic, M., & Hernández-Leo, D. (2021c). Towards Estimating Classroom Orchestration Load using Physiological and Self-Perception Measures. In Hmelo-Silver, C. E., De Wever, B., & Oshima, J. (Eds.), Proceedings of the 14th International Conference on Computer-Supported Collaborative Learning - CSCL 2021 (pp. 287-288). Bochum, Germany: International Society of the Learning Sciences.

Biggs, J. (1999). What the Student Does: teaching for enhanced learning. Higher Education Research & Development, **18**(1), 57–75. https://doi.org/10.1080/0729436990180105

Blikstein, P., & Worsley, M. (2016). Multimodal Learning Analytics and Education Data Mining: using computational technologies to measure complex learning tasks. *Journal of Learning Analytics*, *3*(2), 220-238. https://doi.org/10.18608/jla.2016.32.11

Bloom, B. S. (1956). Taxonomy of educational objectives: Handbook 1. New York: David McKay.

Crespi, F., Amarasinghe, I., Vujovic, M., & Hernández-Leo, D. (2022). Estimating orchestration load in CSCL situations using EDA. *In 2022 international conference on advanced learning technologies (icalt)* (pp. 128-132). IEEE. https://doi.org/10.1109/ICALT55010.2022.00046

Cuendet, S., Bonnard, Q., Do-Lenh, S., & Dillenbourg, P. (2013). Designing augmented reality for the classroom. *Computers & Education*, 68, 557–569. https://doi.org/10.1016/j.compedu.2013.02.015

Dieckmann, M., Hernández-Leo, D., & Amarasinghe, I. (2022). Flagging in teacher-facing orchestration dashboards: factors affecting its use in Pyramid CSCL debriefing. In 2022 International Conference on Advanced Learning Technologies (ICALT) (pp. 133-135). IEEE. https://doi.org/10.1109/ICALT55010.2022.00047

Dillenbourg, P., Zufferey, G., Alavi, H., Jermann, P., Do-Lenh, S., Bonnard, Q., Cuendet, S., & Kaplan, F. (2011). Classroom orchestration: The third circle of usability. In Spada, H., Stahl, G., Miyake, N., & Law, N. (Eds.), Connecting Computer-Supported Collaborative Learning to Policy and Practice: CSCL2011 Conference Proceedings. Volume I — Long Papers (pp. 510-517). Hong Kong, China: International Society of the Learning Sciences. https://doi.dx.org/10.22318/cscl2011.510

Hakami, E., Hakami, L., Hernández-Leo, D., & Amarasinghe, P. I. U. (2023). Triggers of teacher-perceived stressful moments when orchestrating collaborative learning with technology.

Hakami, L., Hernández-Leo, D., Amarasinghe, I., & Sayis, B. (2024). Investigating teacher orchestration load in scripted CSCL: A multimodal data analysis perspective. British Journal of Educational Technology, 55(5), 1926-1949. https://doi.org/10.1111/bjet.13500

Ortega-Arranz, A., Amarasinghe, I., Martínez-Monés, A., Asensio-Pérez, J. I., Dimitriadis, Y., Corrales-Astorgano, M., & Hernández-Leo, D. (2024). Collaborative Activities in Hybrid Learning Environments: Exploring Teacher Orchestration Load and Students' Perceptions. *Computers & Education*, 105105. https://doi.org/10.1016/j.compedu.2024.105105

Prieto, L.P., Sharma, K., Dillenbourg, P. (2015a). Studying Teacher Orchestration Load in Technology-Enhanced Classrooms. In: Conole, G., Klobučar, T., Rensing, C., Konert, J., Lavoué, E. (eds) Design for Teaching and Learning in a Networked World. EC-TEL 2015. Lecture Notes in Computer Science(), vol **9307**. Springer, Cham. https://doi.org/10.1007/978-3-319-24258-3_20

Prieto, L. P., Sharma, K., Kidzinski, Ł., & Dillenbourg, P. (2017). Orchestration load indicators and patterns: In-the-wild studies using mobile eye-tracking. *IEEE Transactions on Learning Technologies*, **11**(2), https://doi.org/10.1109/TLT.2017.2690687

Prieto, L. P., Sharma, K., Wen, Y., & Dillenbourg, P. (2015b). The Burden of Facilitating Collaboration: Towards Estimation of Teacher Orchestration Load using Eye-Tracking Measures In Lindwall, O., Häkkinen, P., Koschman, T. Tchounikine, P. Ludvigsen, S. (Eds.). Exploring the Material Conditions of Learning: The Computer Supported Collaborative Learning (CSCL) Conference 2015, Volume 1. Gothenburg, Sweden: The International Society of the Learning Sciences. https://doi.dx.org/10.22318/cscl2015.206

Salim, R., & Partnerships, D. O. (2020). What is EDA Peak Detection and How Does it Work.

Sharma, K., & Giannakos, M. (2020). Multimodal data capabilities for learning: What can multimodal data tell us about learning?. *British Journal of Educational Technology*, 51(5), 1450-1484. https://doi.org/10.1111/bjet.12993

Taylor, S., Jaques, N., Chen, W., Fedor, S., Sano, A., & Picard, R. (2015). Automatic identification of artifacts in electrodermal activity data. In 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (pp. 1934-1937). IEEE.

van Dooren, M., & Janssen, J. H. (2012). Emotional sweating across the body: Comparing 16 different skin conductance measurement locations. Physiology & behavior, 106(2), 298-304.

Acknowledgments

This work has been supported by PID2023-146692OB-C33 and 2023-26 CEX2021-001195-M granted by MICIU/AEI/10.13039/501100011033. D. Hernández-Leo (Serra Húnter) also acknowledges the support of the ICREA Academia programme.

Appendix A

Table 5
The Stressful Moments Reported by the Teacher in Both Co-Located and Online Learning Settings

Learning Setting	Session No.	The stressful moments reported by the teacher
Co-located	C.1	"I was using flags... it took me some effort to decide what options to flag. There were not really wrong answers in general in this task, which would have been the good ones to flag." (*Actions by students*)
	C.2	"The somewhat stressful moments were when I needed to make decisions about extending or not extending time, and moving to next tasks." (*Time-related issues*) "I also needed to be concentrated when reading the answers and selecting what to flag, but it was good to really pay attention to submissions." (*Actions by students*)
	C.3	"In the submission phase I noticed students were needing extra time just about when the time was finishing. I managed to increase the time on time." (*Time-related issues)* "When students were finished with any phase they started to talk - and the class started to be clearly noisy." (*Actions by students*)
	C.4	"A bit frustrated when I didn't realize to extend even more the submission time" (*Time-related issues*) "The rating time was short and as students asked me about what to do in the rating ... and so many students couldn't complete the rating. I felt bad that I didn't extend the rating time." (*Actions by students*)
	C.5	"I needed to press the "Next level" control several times, until it worked." (*Technological difficulties*)
	C.6	"Not really stressful, it was once difficult to make the "next activity" control work." (*Technological difficulties*)
Online	O.1	"I was running out of time, and needed to reduce time in the activity. However I'm used to this kind of situation, and was not highly stressed." (*Time-related issues*)
	O.2	"I noticed that the scenario for the task that I shared with students was not the one I planned. So I had to read the scenario as well .. while students were completing the activity. In any case I knew all the scenarios very well and it was quick for me to remember it." (*Teacher-related*)
	O.3	No stressful moments were reported.
	O.4	"Not highly stressful, but I didn't notice time was finishing and I increased time at the last second possible." (*Time-related issues*)
	O.5	"Only one moment, prior to the activity - one I noticed I had not published the activity." (*Teacher-related*)
	O.6	"When the initial submission phase finished as I didn't notice - I was not paying attention to time." (*Time-related issues*) "When the time for the first improvement finished - it was before expected in the timeline." (*Time-related issues*)

 © ISLS

International Society of the Learning Sciences

Bridging Group and Individual: *Transitional Spaces* in Collaboration

Xuesong Cang, University of Wisconsin, Madison, xcang@wisc.edu
Michael Tissenbaum, University of Illinois Urbana-Champaign, miketiss@illinois.edu
Sadhana Puntambekar, University of Wisconsin, Madison, spuntambekar@wisc.edu

Abstract: This study investigates how learning traverses between collective and individual levels in collaborative learning by introducing the concept of *Transitional Space*, the temporary and interactional zone where collective knowledge constructions outpace or below individual understanding. Using an approach that integrates students' knowledge complexity levels and participation patterns, we model both group-level dynamics and individual trajectories. Results from two groups reveals that while higher-level engagement often align with increased knowledge at the group level, most individual students within the group deviate significantly from these patterns. These divergence highlights the fragile and contingent nature of transitions, which indicates that students must need a space to reconcile group-generated ideas with their own thinking, the *Transitional Space*. By identifying occurrence of such divergence, this study makes transitions empirically visible. The findings also contribute to theory by challenging assumptions of a spontaneous internalization and providing a methodological bridge between collective discourse and individual learning trajectories.

Introduction

Collaborative learning has been widely recognized as a powerful approach to fostering critical thinking, shared knowledge construction and collaboration skills. Decades of research have illuminated how students build intersubjectivity through group cognition, socially shared regulation, and collective knowledge building (Scardamalia & Bereiter, 2021; Schnaubert & Bodemer, 2019, 2022; Stahl, 2006, 2016). Yet much of this literature implicitly assumes that what is achieved at the group level naturally transfers to the individual: that once students participate in joint meaning-making, they will internalize and extend those insights through the social interactions (Lawrence & Valsiner, 1993; Rogoff, 1995). However, accumulating evidence shows that this assumption does not always hold, pointing a gap in understanding how collective achievements translate into individual learning. For example, empirical studies consistently show uneven benefits across learners, with some students readily transforming collective ideas into personal understanding, while others remain at the periphery of group discourse, rarely express their ideas aligns with the emerging group knowledge construction (Clarke, 2016; Sedova & Navratilova, 2020). The unevenness is particularly salient when considering how individual participation trajectories unfold. Participation in group activities does not guarantee that all students will integrate, reconcile, or deepen the knowledge that emerges collectively (Cang et al., 2024; Puntambekar et al., 2023).

Rather than assuming that group-level engagement guarantees individual uptake, or individual contribution led to group work improvement, we argue that learners must move intentionally between social participation and individual sense-making (Damșa & Ludvigsen, 2016; Järvelä & Hadwin, 2013; Olsen et al., 2021). And this movement is contingent, fragile, and design dependent, and it is reciprocal in that collaboration both supports and demands conscious individual engagement. This paper examines how learning moves across levels, between the group and individuals. We introduce the concept of the *Transitional Space*: the discursive and cognitive zone in which students reconcile group knowledge with their own developing understandings. And we investigate how this *Transitional Space* emerges and can be identified in technology-enhanced collaborative science learning, and how individual trajectories vary in traversing it. To do this, it requires us to investigate the group collaborative patterns and individual's understanding trajectories. Therefore, the research questions are:

1. What are the two group's participation trajectories? How does it imply
2. What are the individual student's trajectories in the group collaboration?
3. How does the alignment or misalignment between individuals' learning trajectories and group progress during collaboration indicate the emergence of *Transitional Space*?

Theoretical perspective

Bridging collective and individual learning

 © ISLS

Collaborative learning theory, particularly sociocultural and knowledge-building traditions, has richly described how shared meaning-making unfolds at the group level (Scardamalia & Bereiter, 2021; Stahl, 2006). Mutual regulation, group awareness and group cognition are treated as central mechanisms of productive collaboration (Järvelä & Hadwin, 2013; Schnaubert & Bodemer, 2022). However, this orientation has often privileged collective outcomes, leaving the individual's uptake of group-generated knowledge less explained. While other studies, especially cognitive perspective grounded, highlight individual's change by treating the collaboration as context. This approach makes it difficult to explain how specific features of group interaction shapes individual learning mechanism of group collaboration. Neither group-focused nor individual-centered approach fully reveal the learning traverse levels adequately. Instead of treating individual and collective learning as two units that can be analyzed exclusively, they must be brought together (Hakkarainen, 2009; Stahl, 2013).

Vygotskian perspectives emphasize that internalization emerges from participation in social interaction (Lawrence & Valsiner, 1993; Rogoff, 1995). Yet internalization of collective level's ideas is not spontaneous and effortless, especially when new ideas conflict with individual's existing understanding. Leaners must bridge the space between public discourse and private understanding (Matusov, 1998; Wells, 2015). This transition varies across individuals and contexts, shaped by differences in prior knowledge, interactional positioning, and task design (Järvelä et al., 2015). The challenge is therefore to identify when this transition between levels emerges and how to scaffold students' understanding when it emerges (Damșa, 2014; Olsen et al., 2021). It indicates the need of a conceptual lens that can account for transition between levels, when collective knowledge becomes available or unavailable for individual sense-making. This study proposes the concept of *Transitional Space*, identifying when transitions between levels occurs and explain how it is enacted.

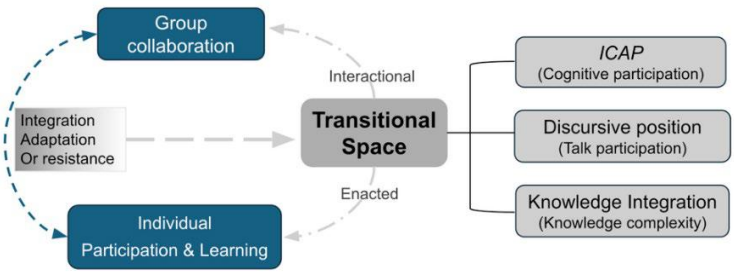
Transitional Space and its operationalization

Conceptualization of *Transitional Space*

The *Transitional Space* is when individuals engage with group ideas generated in collaboration, either integrating, adapting, or resisting them to refine students' own thinking as students are in the discourse of collaborative work (Figure 1). As a result, individual's learning can move across levels, from individual to collective or vice versa, in the collaborative activity (Stahl, 2013). In other words, it is a temporary and interaction-dependent zone in collaborative learning where group-level ideas become available for individual uptake (White, 2018). It is theorized as a construct linking group and individual learning with two key features.

First, it is *enacted* because it is not pre-planned or structured in advance, but emerges dynamically in interaction, specifically when group achieve understanding advancement, or higher order of reasoning that go beyond any single participant's prior understanding (Roschelle & Teasley, 1995; Stahl, 2016). As such, a zone where individual might take up new ideas, emerges through the group discourse. But whether that happens depends on their ability to engage, process and internalize what the group has constructed (Isohätälä et al., 2020). Second, it is *interactional* because the existence and accessibility of this zone depend on how learners interact one another, including how they interpret and build on each other's ideas. The discursive moves, like talk enabling joint sensemaking, and shared artifacts that externalize and stabilize collective thinking are all possible ways to interactions (Hakkarainen & Seitamaa-Hakkarainen, 2022; Suthers, 2006). Yet, access to this space is uneven. Some learners actively enter and engage with group ideas while others remain peripheral, missing the opportunity to reconcile their own thinking with the group's (Sedova & Navratilova, 2020).

Figure 1
Transitional Space



This points it out that *Transitional Space* is not guaranteed, but a fragile and contingent phenomenon that arises under specific discursive and social conditions, and usually at the unpredicted way (Borge & Mercier, 2019). Moreover, transitions may occur rapidly and are often not directly observable as discrete events. To trace collaborative trajectories and identify instance of transitions across levels requires us to focus on the patterns that

become visible or can be inferred through shifts in participation, discourse and artifacts. For this reason, understanding how it operates requires attention not only to the quality of group discourse, but also to the trajectories of individual participation (Dornfeld et al., 2017). That is, how different learners engage with, contribute to, and integrate ideas (Roschelle, 2012). We operationalize it through two dimensions, knowledge complex levels and participation by using three measurements: *ICAP*, discursive position and *Knowledge Integration* (KI) (Chi & Wylie, 2014; Linn et al., 2013; Liu et al., 2008).

Identifying *Transitional Space:* A trace of learning and participating trajectories

First, the *ICAP* (Chi & Wylie, 2014) is used to capture differences in cognitive participation, including passive, active, constructive, and interactive, when individuals participate in group work. It not only reveals whether students are present in the discourse, but also the level at which they are processing ideas and generating understanding beyond what is explicitly presented. This dimension goes beyond mere talk, unpacking how students engage with collective reasoning and higher-order thinking may be occurring. Second, *Discursive Position* emphasizes the social and interactional aspect of learners in group talk (Azmitia & Montgomery, 1993; Clarke, 2016). Students vary in how they contribute, from private self-talk, to responding when promoted, to transactive contribution and offering unsolicited talk that co-construct knowledge (Diez-Palomar et al., 2021; Kumpulainen & Rajala, 2017). By analyzing the language in use, we capture how learners negotiate ideas, respond to peers and shape the group's reasoning collectively (Tao & Zhang, 2021; Tong & Chan, 2023). Both aspects capture the participations. Third, *Knowledge Integration* (*KI*) traces understanding by examining the content and structure of ideas that learners articulate (Liu et al., 2008). *KI* identifies how students connect scientific concepts, from fragmented or minimally integrated knowledge to complex, coherent integration.

Taken together, it provides a multidimensional view of how transitional space is enacted. Tracing trajectories of learning and participation within collaborative activities reveals how students move from shared group reasoning toward their own understanding and highlights emergent of transitions. Thus, it indicates whether and when a space for reconciling ideas from others is needed in collaborative learning.

Methods

Context

8th-grade students from Midwestern U.S. classes participated in a three-week biology unit on designing community gardens. The unit was taught by the same teacher who got professional training before implementing the unit. In home groups of two to four, students completed 15 activities (Table 1) with teacher's facilitation. Students started with practicing the use of technological tools, followed by the topic introduction, exploring *Why Plants Are Important* and conducting food research, and progressed through experiment on the biotic & abiotic factors, hypothesis theorizing and testing, and analyzing data, and making claims of different plants, starting from the tomato plants and continuing to investigate other plants (e.g., lettuce, watermelon, bell peppers). The activities aimed to develop science skills and deepen understanding of key concepts.

The activities were supported by a technology-enhanced learning environment (Tools column in Table 1). While each activity primarily relied on a specific tool, students could access others as needed. For instance, students conducted experiments using the plant simulation tool (PlantSim) (Sultana, et al., 2024) and engaged in science reasoning (e.g., making hypotheses, synthesizing findings, and drawing inferences) using the Idea Wall and/or the online digital science notebook (DigiNb).The Idea Wall, an interactive science reasoning tool, enables students to externalize and synchronize their ideas, allowing group members to create, view, and manipulate shared notes (Zeng et al., 2024). DigiNb, designed to integrate various curriculum units, includes a *Notes* function (Notes) for jotting down ideas, which can later be revisited or transferred to the Idea Wall. Both the Idea Wall and PlantSim are embedded within DigiNb, providing a cohesive digital learning environment. Students accessed all tools by logging into their accounts on their devices.

Table 1
Activities and tools

Activity (a)	Day(d)	Tools
a1: practice using tools	d2	Idea Wall
a2: research on "Why plants are important?"	d3	Notes
a3: conduct food research	d3&4	Notes
a4: share out food research finding	d4&5	Idea Wall
a5: explore biotic and abiotic factors	d5	Cards (Non-digital tool) *
a6: explore dependent & independent variables	d6	Notes

a7: make hypothesis for tomato plants	d7	Idea Wall
a8: experimented on tomato plants	d7	Plant Simulation (PlantSim)
a9: analyze tomato plants	d8	Online Digital Notebook (DigiNb)
a10: synthesize findings of tomato plants	d9	Idea Wall
a11: make hypothesis for other plants	d9	Online Digital Notebook (DigiNb)
a12: experimented on other plants	d9	Plant Simulation (PlantSim)
a13: analyze other plants	d10	Online Digital Notebook (DigiNb)
a14: plan the design of the community garden	d11	Idea Wall
a15: scratch final design	d12	Paper Poster (Non-digital tool) *

Notes: ***1***. Letter "a" is the activity, e.g., "a1" is activity 1; ***2***. Letter "d" is the day, e.g., "d3&4" means students started activity on day 3 and extended it to day 4; ***3.*** Four types of technological tools and two non-technological tools (marked by *) are available; ***4.*** Cards with different images denoting biotic or abiotic factors, and paper poster was used to sketch community garden design; ***5.*** Day 1 is not included when students did the pre-test.

Participants and data sources

Seven students worked in two groups' (N_1 = 4, N_2 = 3) but from two classes consented to be recorded. Group 1 consisted of two male and two female students, and Group 2 consisted of three female students. Three of them are whites and four of them non-whites. They were included in this study because both groups used all (digital) tools throughout the curriculum unit, and with minimum class absence. Their participation in activities was video-recorded, and their device interactions were screen-recorded. Videos were organized chronologically according to the classroom activities implemented across curriculum unit, totaling 30 transcriptions. Students finished some of the activities in more than one classes (45 mins per class) because of the practical reasons, such as additional time needed for certain groups, or teacher needed to facilities students to do some of other tasks that were regular school activities. Videos were transcribed verbatim. Tool interactions, including (1) the written ideas within the notes, online digital notebook, and (2) manipulation of notes in Ideal Wall (e.g., move notes), provided supplementary insights into students' idea externalizations beyond verbal contributions. Transcriptions were organized by turns of talk, contextualized by preceding and following turns to capture interaction flow. Teacher interactions, as facilitative rather than directive, were excluded because these interactions did not reflect students' participation in group work.

Coding

We used discourse analysis techniques, inductively and deductively (Derry et al., 2010; Jordan & Henderson, 1995), to examine students' collaborative group work through three key dimensions. First, we adapted the ICAP framework (Chi & Wiley, 2014) to analyze students' participation in group activities, based on both the conversations in the group and the tool interaction. Their participation was categorized as passive, active, constructive, or interactive (Table 2), representing increasing degrees of contribution and collaboration. It allows us to assess the extent to which they contributed to the collaboration and engaged with peers. *Passive participation*, the lowest level, referred to instances where students merely received information or provided limited, unelaborated responses. *Active participation* reflected a higher level, where students performed meaningful actions that supported the group's progress. *Constructive participation* indicated an even higher level, when students generated ideas that went beyond or reinterpreted shared information, *Interactive participation*, the highest level, captured moments when students built upon and evaluated one another's ideas to jointly develop more sophisticated understanding.

Table 2

Interactive-Constructive-Active-Passive (ICAP) Taxonomy for Group Collaborations

ICAP	Descriptions	Examples
4-Interactive	Develop knowledge and science practices, becoming evaluative via dialogue characterized by frequent turn-taking, ideas mutual exchange, and primarily above constructive	• Argue a position or make claims using diverse views • Building on others' contributions • Critique to build shared knowledge • Collaboratively taking notes based on inferences • Arguing scientific practices
3-Constructive	Generate new ideas that go beyond the information given or ideas shared	• Ask to clarify or respond to this ask • Adapt peers' ideas • Generate notes to synthesize findings

 © ISLS

2-Active	Take meaningful actions (verbally or physically) to develop understanding	• Respond to peers' tool manipulation • State observations from experiments • Propose ideas/actions to achieve goals
1-Passive	Receiving information passively, externalizing ideas, or maintaining focused attention	• Respond when asked, sometimes meaninglessly • Restate content from materials/ notes • Respond mechanically to suggestions

Second, we examined students' discursive position by categorizing their talk into four hierarchical types: unsolicited talk, transactive talk, solicited talk, and self-talk (Azmitia & Montgomery, 1993; Clarke, 2016) (Table 3). *Unsolicited talk* occurs when students voluntarily share ideas, often expanding group knowledge beyond its initial scope. *Transactive talk* involves the frequent exchange and negotiation of ideas. *Solicited talk* reflects responses to others' prompts aiming to build mutual understanding, while *self-talk* consists of self-directed remarks or responses to prompts with no intention to share ideas. Analyzing these patterns provided insight into how students externalized thoughts and co-constructed understanding collectively.

Table 3
Discursive Position (DP)

DP	**Descriptions**	**Examples**
4-Unsolicited talk	Interlocutors engage in unprompted talk, often leading to knowledge co-construction beyond the sum of individual ideas	• Elicit others to involve in conservation • Lead conversation by asking peers' ideas • Synthesize ideas to get solutions or resolve conflicts
3-Transactive talk	Interlocutors share sources & messages simultaneously and/or acknowledge/deny/assimilate ideas	• Justify or elaborate on ideas/actions • Request justification, elaboration, or clarification spontaneously
2-Solicited talk	Interlocutors involve mechanical responses to procedural questions, not knowledge construction.	• Respond to peers' tool manipulation • State observations from experiments • Propose ideas/actions to achieve goals
1-Self-talk	Private talking or monologizing w/o the intention to get response from the peers	• Read out written notes of them owns • Repeat ideas from learning resources • Narrate one's own actions when working

Last, we analyzed knowledge advancement using the *KI* framework (Liu et al., 2008). It assesses students' understanding by categorizing their connections between concepts. The highest level is the *complex link* when three or more concepts are connected correctly, *full-link* as next level if two concepts are explicated connected correctly, *partial-link* as an even lower level of understanding by identifying potential relationships, and *no link* is the lowest level when invalid/wrong connections occur (more details in our previous studies).

Together, three-dimensions allows to capture the nuances of students' participation and conceptual understanding in collaborative activities. Two raters independently coded 17% of the dataset, achieving an inter-rater reliability of .95, .93 and .94 for *ICAP*, Discursive and *KI* through iteratively refined the coding scheme. All disagreements were resolved through discussions, and the first author coded the remaining data.

Analysis

To explore the relationship between group-level participation and knowledge integration, we constructed a smoothed scatterplot using regression, modeling *KI* as a function of Discursive Position. *ICAP* level was encoded as a third dimension using a continuous color gradient. This visualization (Graph 1) allowed us to see the general patterns in how higher-order *KI* emerged in relation to the discursive and participation of the group. Although non-parametric, this fit served as an exploratory model guiding the subsequence analysis. Next, we further examine if two groups show difference statistically in their participation patterns. We conducted a cross-group model comparison. Specially, we tested whether the effect of Discursive position, *ICAP* and their interaction on *KI* varied significantly between Group 1 and Group 2. Two linear models were constructed. The baseline model predicted *KI* from Discursive Participation, ICAP and their interaction (Group baseline model: *KI* ~ Discursive x *ICAP*), assuming that the relationship was identical across both groups. The extended group model added interaction terms with the group variable (Extended group baseline model: *KI* ~ Discursive x *ICAP* x Group), allowing the relationship between Discursive position and *ICAP* to vary by groups.

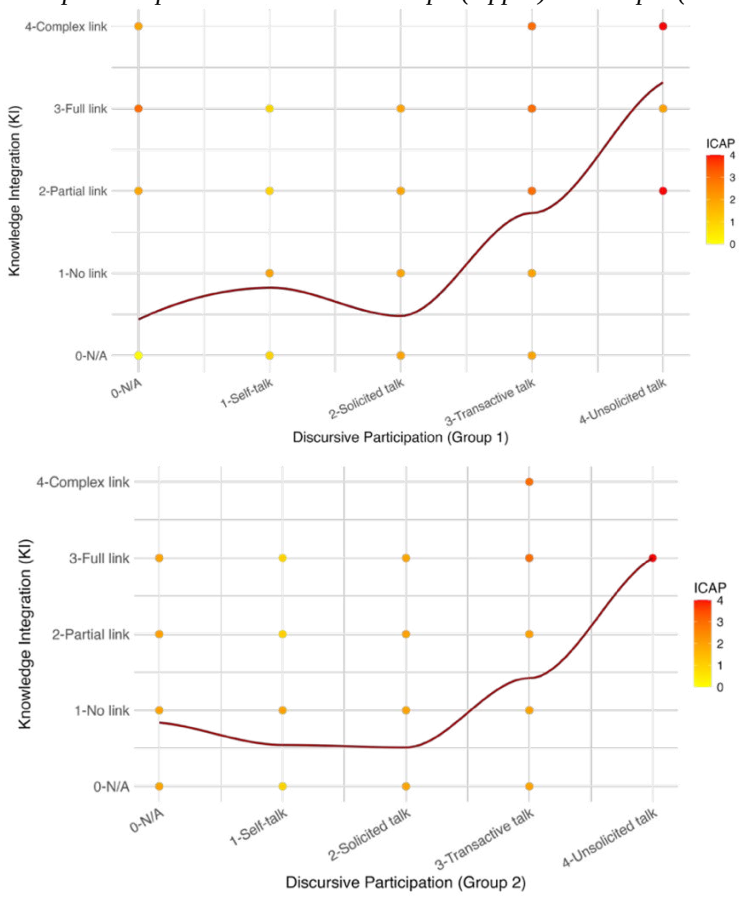
Last, we investigated whether individual students' learning trajectories deviated from the overall group pattern. Similarly, we built the smoothed scatterplot for each focal student. After, we did a nested model comparison using the group level regression as the baseline (*KI* ~ Discursive x *ICAP*). We next conducted extended models by adding an interaction with a dummy-coded variable that distinguished the focal student from all other group members (Table 4). For example, in Group 1, a dummy variable was created such that any contributions from student AR were coded as 1, and contributions from other group members were coded as 0, when we compared AR to group pattern. Each baseline-extended model pair compared using *ANOVA*. A statistically significant improvement in the model fit ($p < .05$) indicated that the focal student's participation trajectory differs from the group level. A non-significant result suggested alignment with the group trajectory.

Results

To understand how *Transitional Space* emerges in collaborative learning, we examined dimensions of participation: *ICAP* levels and Discursive Participation, along with *Knowledge Integration*. We first mapped group-level patterns using smoothed scatterplots and conducted the regression models to identify when and how higher-order integration occurred during collaboration. We then examined individual trajectories to access whether students aligned with or diverged from group pattern. Using visualization and nested model comparison, we evaluated how individuals engage with group discourse, offering insight into when *Transitional Space* is enacted as students work to reconcile collective ideas with personal understandings.

Graph 1

Group Participation and KI Plot: Group1 (Upper) & Group 2 (lower)



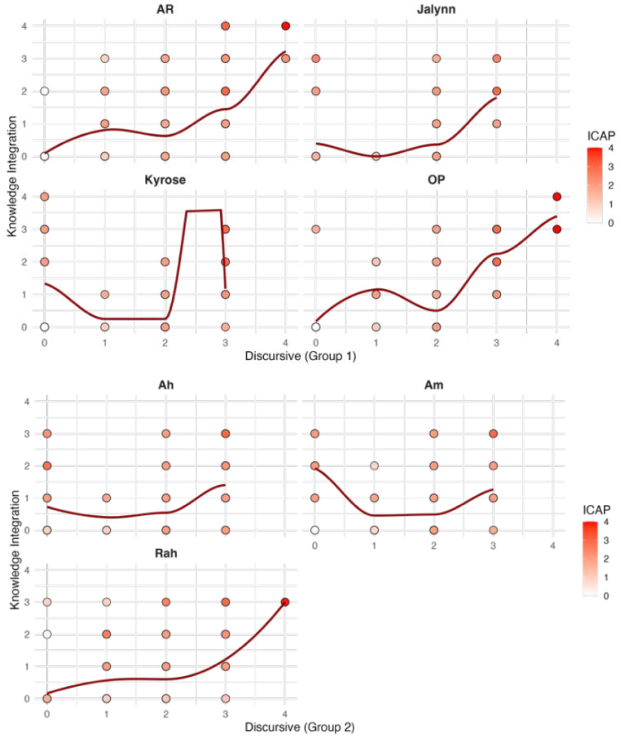
At the group level, we observed from both groups that knowledge integration tends to converge when learners participated the collaboration constructively or interactively and generate transactive and unsolicited talk. And their participation trajectory showed visually similar pattern (Graph 1). However, the model comparison revealed a statistically significant improvement in fit for the extended model: $F_{(63, 2235)} = 3.21$, $p < .001$. This result indicates that the relationship between Discursive Position, *ICAP* and *KI* was not consistent between groups. Instead, the dynamics of how participation, including both discursive and *ICAP*, contributed to *KI* varied significantly between Group 1 and 2. In other words, the patterns of productive participation leading to KI were

group dependent, reinforcing the importance of contextual and interactional differences in collaborative environments. It also further explains that transitions are enacted dynamically.

At the individual level, however, we also observe that each student's participation within one group mostly diverge from the group participation patterns (Graph 2). We test if one of the student's participation patterns differ from group participation by compare student's individual participation model to the group participation model (overall results in Table 4). In Group 1, we observed that adding a "AR vs. Others" contrast significantly improves the model ($p < 0.001$). This means students AR's learning trajectory is statistically different from the group-level trajectory (the combined pattern of all other students). Similarly, adding an "Kyrose vs. Others" and an "Jalynn vs. Others" contrast greatly improves the model at $p < .01$. But adding an "OP vs. Others" contrast does not improve the model ($p = 0.58$).This means OP's learning trajectory is not significantly different from the group-level trajectory. In conclusion, only OP's participation and KI align with the group level's three dimensions. All others are statistically different from the group level.

Graph 2

Individual Student' Participation Trajectory: Group1 (upper) and Group2 (lower)



We also tested for Group 2 and observed that adding a "Rah vs. Others" contrast and "Am vs. Others" contract significantly improves the model at $p < 0.001$. This means Rah's and Am's learning trajectory is statistically different from the group-level trajectory (the combined pattern of all other students). However, adding an "Ah vs. Others" contrast does not improve the model ($p = 0.63$).This means Ah's learning trajectory is not significantly different from the group-level trajectory. In conclusion, only Ah's participation and KI align with the group level's three dimensions. All others are statistically different from the group level.

Table 4

Nested Model Comparisons of Individual vs. Group Trajectories

Group	Baseline Model	Extended Model	*F*	*p*-value
1	KI ~ Discursive × ICAP	KI ~ Discursive × ICAP × (AR vs. Others)	12.88 *	< .001
		KI ~ Discursive × ICAP × (Kyrose vs. Others)	27.39 *	< .01
		KI ~ Discursive × ICAP × (Jalynn vs. Others)	5.54 *	< .01
		KI ~ Discursive × ICAP × (OP vs. Others)	.72	.58
2	KI ~ T Discursive × ICAP	KI ~ Discursive × ICAP × (Rah vs. Others)	7.71 *	< .001
		KI ~ Discursive × ICAP × (Am vs. Others)	16.29 *	< .001
		KI ~ Discursive × ICAP × (Ah vs. Others)	.65	.63

 © ISLS

Note: Baseline models predict *KI* from Discursive position, *ICAP*, and their interaction, assuming all students share the same trajectory. Extended models add a focal student's interaction term to test whether an individual student's trajectory differs significantly from the group-level trajectory. Reported *p*-values are from *ANOVA* comparisons of extended vs. baseline models.

Discussion

This study aims to examine how learning traverses between group and individual levels in collaborative learning, aiming to make the concept of *Transitional Space* empirically observable and identifiable. While smoothed scatterplot showed visual similarities in group participation patterns, further statistical analysis revealed that collaborative dynamics varied substantially across groups. The way students engaged in discursive participation, showed *ICAP* levels, and *KI* was not uniform, even when surface-level patterns appeared similar. This variability reinforces the idea that collaborative learning is a highly contextual process, shaped by group composition, task structure, and discourse, rather than a fixed or prescriptive process (Damșa, 2014; Puntambekar et al., 2023). The converge of *ICAP* and *KI* at higher levels, reflected in increased knowledge complexity (Linn et al., 2013; Roschelle, 2012), offers insights into when transitions are more likely to occur spontaneously. Although transitions cannot be directly observed, they can be inferred from pattens in discourse and participation across levels. This also suggests that collective progress is driven by continuous individual contributions, as learners more fluently externalize their thinking to the group.

Beyond these group-level differences, we found strong evidence of divergence between individual and group learning trajectories. Nested model comparisons showed that in both groups, only one student's trajectory aligned with group-level pattern, while most students significantly deviated. This consistent result suggests that group level advancement and success does not necessarily represent the individual learning of its members. While group activity may provide a collective trajectory toward knowledge integration, individual leaners followed distinct paths, sometimes aligned, but often misaligned with the group's progression (Dornfeld et al., 2017; White, 2018). This indicates that individual sense-making derived from collaborative work requires additional support in order to enhance the benefits that collaboration provides for each learner.

These misalignment further points to the importance of *Transitional Space*, the temporary and contingent zone that emerges when group-level knowledge advancement and participation outpace individual understanding. In such moments, group discourse may generate higher-order reasoning or improvement that exceed the current grasp of some members (Roschelle & Teasley, 1995; Stahl, 2016). *Transitional Space* is not a pre-structured feature of the learning environment, but an enacted and interactional phenomenon, brough into being through discourse, shared artifacts and uneven participation (Suthers, 2006; Hakkarainen, 2009; Damșa & Ludvigsen, 2016). Critically, it is fragile. Its emergence depends on the discursive and social conditions that allow individual to revisit, reinterpret and rework group ideas in ways that become personally meaningful. Identifying when and how transitional space arises, therefore, requires examine both group-level discourse and the individual trajectories within it (Cang et al., 2024). This can further assist to improve the quality of collaborative work and individual learning.

This study offers insights on identifying transitions as an empirical phenomenon, rather than a purely theoretical one. Research in sociocultural learning and knowledge-building traditions has implicitly draw upon internalization but lack of the tools for detecting when and how individual uptake from group progress occurs (Scardamalia & Bereiter, 2021; Lawrence & Valsiner, 1993; Stahl & Hakkarainen, 2021). This study provides a concrete approach to trace when collaborative progress may successfully or unsuccessfully translate into individual understanding. In doing so, this study contributes theoretically by challenging the assumption that internalization is an inevitable or uniform outcome of social interactions. Instead, it foregrounds the unevenness of learning in group settings and identifies the *Transitional Space* as a critical arena where learning can be supported. While *Transitional Space* cannot be guaranteed, understanding its dynamics offers new insights into how learning environment might be designed to support individual sensemaking within collective activity.

Limitation and future work

Although this study provides insights on identifying transitional space as a contingent and fragile dimensions of collaborative learning, it did not examine how it varied by activity or tools. Future work will expand it with microanalysis to capture how reconciliation unfolds in interaction, and to better understand the role of specific instructional supports and temporal rhythms, such as tool shifts, reflection chance, in enabling *Transitional Space*. Further. while higher proportions of integrative moves suggest greater opportunities for individual reconciliation, they do not directly confirm that learners internalized or retained group-generated ideas. Longitudinal or transfer measures would help to understand sustained individual uptake. Finally, with only two groups working within a particular science curriculum unit may limit the generalizability.

 © ISLS

References

Azmitia, M., & Montgomery, R. (1993). Friendship, transactive dialogues, and the development of scientific reasoning. *Social Development, 2*(3), 202–221.

Borge, M., & Mercier, E. (2019). Towards a micro-ecological approach to CSCL. *International Journal of Computer-Supported Collaborative Learning, 14*, 219–235.

Cang, X., Zeng, L., Tissenbaum, M., Gnesdilow, D., & Puntambekar, S. (2024. June). Convergent Group Understanding: Indicator of Knowledge Integration?. In *Proceedings of the 17th International Conference on Computer-Supported Collaborative Learning-CSCL 2024, pp. 197-200*. International Society of the Learning Sciences.

Chi, M. T., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist, 49*(4), 219–243.

Clarke, S. N. (2016). Student agency to participate in dialogic science discussions. *Learning, Culture and Social Interaction, 10*, 27–39.

Damşa, C. I. (2014). The multi-layered nature of small-group learning: Productive interactions in object-oriented collaboration. *International Journal of Computer-Supported Collaborative Learning, 9*, 247–281.

Damşa, C. I., & Ludvigsen, S. (2016). Learning through interaction and co-construction of knowledge objects in teacher education. *Learning, Culture and Social Interaction, 11*, 1–18.

Derry, S. J., Pea, R. D., Barron, B., Engle, R. A., Erickson, F., Goldman, R., ... & Sherin, B. L. (2010). Conducting video research in the learning sciences: Guidance on selection, analysis, technology, and ethics. *The Journal of the Learning Sciences, 19*(1), 3–53.

Diez-Palomar, J., Chan, M. C. E., Clarke, D., & Padros, M. (2021). How does dialogical talk promote student learning during small group work? An exploratory study. *Learning, Culture and Social Interaction, 30*, 100540.

Dornfeld, C., Zhao, N., & Puntambekar, S. (2017). A mixed-methods approach for studying collaborative learning processes at individual and group levels. Philadelphia, PA: *International Society of the Learning Sciences*.

Hakkarainen, K. (2009). A knowledge-practice perspective on technology-mediated learning. International *Journal of Computer-Supported Collaborative Learning, 4*, 213–231.

Hakkarainen, K., & Seitamaa-Hakkarainen, P. (2022). Learning by inventing: Theoretical foundations. In Invention pedagogy – *The Finnish approach to Maker education* (pp. 15–27). Routledge.

Isohätälä, J., Näykki, P., & Järvelä, S. (2020). Convergences of joint, positive interactions and regulation in collaborative learning. *Small Group Research, 51*(2), 229–264.

Järvelä, S., & Hadwin, A. F. (2013). New frontiers: Regulating learning in CSCL. *Educational Psychologist, 48*(1), 25–39.

Järvelä, S., Kirschner, P. A., Panadero, E., Malmberg, J., Phielix, C., Jaspers, J., ... & Järvenoja, H. (2015). Enhancing socially shared regulation in collaborative learning groups: Designing for CSCL regulation tools. *Educational Technology Research and Development, 63*, 125–142.

Jordan, B., & Henderson, A. (1995). Interaction analysis: Foundations and practice. *The Journal of the Learning Sciences, 4*(1), 39–103.

Kumpulainen, K., & Rajala, A. (2017). Dialogic teaching and students' discursive identity negotiation in the learning of science. *Learning and Instruction, 48*, 23–31.

Lawrence, J. A., & Valsiner, J. (1993). Conceptual roots of internalization: From transmission to transformation. *Human Development, 36*(3), 150–167.

Linn, M. C., Eylon, B. S., & Davis, E. A. (2013). The knowledge integration perspective on learning. In *Internet Environments for Science Education* (pp. 57–74). Routledge.

Liu, O. L., Lee, H. S., Hofstetter, C., & Linn, M. C. (2008). Assessing knowledge integration in science: Construct, measures, and evidence. *Educational Assessment, 13*(1), 33–55.

Matusov, E. (1998). When solo activity is not privileged: Participation and internalization models of development. *Human Development, 41*(5–6), 326–349.

Olsen, J. K., Rummel, N., & Aleven, V. (2021). Designing for the co-orchestration of social transitions between individual, small-group and whole-class learning in the classroom. *International Journal of Artificial Intelligence in Education, 31*(1), 24–56.

Puntambekar, S., Gnesdilow, D., & Yavuz, S. (2023). Understanding the effect of differences in prior knowledge on middle school students' collaborative interactions and learning. *International Journal of Computer-Supported Collaborative Learning, 18*(4), 531–573.

Rogoff, B. (1995). Observing sociocultural activity on three planes: Participatory appropriation, guided participation, and apprenticeship. In J. V. Wertsch, P. del Río, & A. Alvarez (Eds.), *Sociocultural studies of mind* (pp. 139–164). Cambridge University Press.

Roschelle, J. (2012). Learning by collaborating: Convergent conceptual change. In *CSCL* (pp. 209–248). Routledge.

Roschelle, J., & Teasley, S. D. (1995). The construction of shared knowledge in collaborative problem solving. In *Computer Supported Collaborative Learning* (pp. 69–97). Springer Berlin Heidelberg.

Scardamalia, M., & Bereiter, C. (2021). Knowledge building: Advancing the state of community knowledge. In *International Handbook of Computer-Supported Collaborative Learning* (pp. 261–279). Springer.

Schnaubert, L., & Bodemer, D. (2019). Providing different types of group awareness information to guide collaborative learning. *International Journal of Computer-Supported Collaborative Learning, 14*, 7–51.

Schnaubert, L., & Bodemer, D. (2022). Group awareness and regulation in computer-supported collaborative learning. *International Journal of Computer-Supported Collaborative Learning, 17*(1), 11–38.

Sedova, K., & Navratilova, J. (2020). Silent students and the patterns of their participation in classroom talk. *Journal of the Learning Sciences, 29*(4–5), 681–716.

Stahl, G. (2006). *Group cognition: Computer support for building collaborative knowledge*. MIT Press.

Stahl, G. (2013). Learning across levels. *Computer Supported Learning*, 8, 1–12. https://doi.org/10.1007/s11412-013-9169-0

Stahl, G. (2016). From intersubjectivity to group cognition. *Computer Supported Cooperative Work (CSCW)*, 25, 355–384.

Sultana, A., Zeng, L., Wang, M., Cernova, S., Cang, X., Gnesdilow, D., ... & Mazalek, A. (2024). A plant simulation tool for collaborative biology experiments in middle-school classrooms: An in-the-wild study. In *Proceedings of the 50th Graphics Interface Conference* (pp. 1-12).

Suthers, D. D. (2006). Technology affordances for intersubjective meaning making: A research agenda for CSCL. *International Journal of Computer-Supported Collaborative Learning, 1*, 315–337.

Tao, D., & Zhang, J. (2021). Agency to transform: How did a grade 5 community co-configure dynamic knowledge building practices in a yearlong science inquiry? *International Journal of Computer-Supported Collaborative Learning, 16*, 403–434.

Tong, Y., & Chan, C. K. (2023). Promoting knowledge building through meta-discourse and epistemic discourse understanding. *International Journal of Computer-Supported Collaborative Learning, 18*(3), 353–391.

White, T. (2018). Connecting levels of activity with classroom network technology. *International Journal of Computer-Supported Collaborative Learning, 13*, 93–122.

Wells, G. (2015). Dialogic learning: Talking our way into understanding. In *Education as Social Construction* (pp. 62–84).

Zeng, L., Cang, X., Sultana, A., Cernova, S., Wang, M., Gnesdilow, D., ... & Tissenbaum, M. (2024, June). Spatial manipulative note-taking tool for small group face-to-Face collaboration in science class. *Proceedings of the 17th International Conference on Computer-Supported Collaborative Learning-CSCL 2024*.

Acknowledgments

We thank the students and the teacher who participated in this study. The research reported in this study is supported by a grant from the National Science Foundation (2010357) to the second and last authors

International Society of the Learning Sciences

When Peers Outperform AI (and When They Don't): Interaction Quality Over Modality

Caitlin Morris, MIT Media Lab, camorris@media.mit.edu
Pattie Maes, MIT Media Lab, pattie@media.mit.edu

Abstract: As AI increasingly enters the classroom, what changes when students collaborate with algorithms instead of peers? We analyzed 36 undergraduate students learning graph theory through peer collaboration (n=24) or AI assistance (n=12), using discourse analysis to identify interaction patterns shaping learning outcomes. Results suggest a collaboration quality divide: high-quality peer interactions generated curiosity and engagement that AI couldn't match, yet low-quality peer interactions performed worse than AI across dimensions. AI showed a paradoxical pattern, building confidence in knowledge while reducing curiosity and deeper engagement. Interaction quality emerged from dynamic patterns rather than individual traits, with early discourse markers predicting outcomes. Students treated AI as a transactional information source despite its collaborative design, revealing fundamental differences in human versus algorithmic engagement. Our findings suggest AI in education need not replace peer learning but can recognize struggle and support both peer and AI interactions toward productive learning experiences.

Introduction

Educational technology adoption has accelerated dramatically, with AI-powered tutoring systems now reaching millions of students worldwide. As AI assistants become more sophisticated, they increasingly substitute for peer interaction in educational settings. This rapid integration raises fundamental questions about how AI changes the nature of learning, particularly the social dimensions that have long been central to education. These questions take on urgency against a backdrop of rising social disconnection among young people (Cacioppo et al., 2018; Pentina et al., 2023; Twenge et al., 2021).

Vygotsky's foundational work on the Zone of Proximal Development emphasized that learning is inherently social, occurring through interaction with more- and less- knowledgeable others who provide scaffolding within the learner's developmental range (Vygotsky, 1978). Decades of research demonstrate that peer collaboration enhances learning through multiple mechanisms that extend beyond information exchange. When learners explain concepts to peers, they improve their understanding, build secondary skills, and strengthen retention; this phenomenon is extensively documented in learning sciences research (Chi et al., 2001; Roscoe & Chi, 2007). Socially shared regulation, where groups collectively monitor and adjust their learning strategies, predicts both learning process quality and learning outcomes in collaborative settings (Järvelä & Hadwin, 2013; Malmberg et al., 2015). These social processes foster curiosity, motivation, and deeper engagement with material, extending beyond material transfer. Critically, they depend on the gap between two perspectives: the productive friction of navigating uncertainty with another mind that has its own incomplete understanding, strategies to regulate, and reasoning that can surprise. This raises a question prior work has not directly addressed: when an AI replaces a human peer, which of these mechanisms can still operate? This theoretical question motivates our focus on interaction process rather than outcomes alone, and on curiosity expressions, discourse balance, and explanation patterns as behavioral signatures of these mechanisms.

Consider two scenarios from our research. In one, two students puzzle over an applied graph theory problem together, suggesting half-formed ideas, speaking hesitantly, until one suggests, "Wait, what if we draw it like a map first?" This spark of understanding ripples through their conversation, leading to deeper exploration together. In another scenario, a student works with an AI assistant that probes the student for their own understanding, while also guiding with clear, accurate explanations of centrality measures and weighting. The student gains greater confidence in the terminology but never ventures beyond the assigned problems. Which scenario better serves learning? Our results suggest that the answer is neither simple nor universal.

This study examines these questions through detailed discourse analysis of collaborative learning sessions, comparing peer-peer and human-AI interactions. We analyzed 36 undergraduate students learning graph theory, using computational methods to identify specific discourse patterns that shape learning experiences. We identify preliminary evidence of a collaboration quality divide that challenges assumptions about both modalities: while high-quality peer interactions generate curiosity and engagement that AI cannot match, low-quality peer collaborations perform worse than AI across multiple dimensions. Even identical AI behavior produces divergent

 © ISLS

outcomes depending on how students engage with it, suggesting that interaction quality emerges from dynamic patterns rather than the capabilities of either party alone.

These patterns appear early enough to enable real-time intervention, pointing toward a future where one role of AI is not to replace peer collaboration but to detect and support it when needed. By identifying which collaborations need support and understanding the complementary affordances of peers (sparking curiosity, generating interest) and AI (consistent knowledge delivery, procedural confidence), we offer grounding for designing systems that preserve the best of social learning while leveraging AI's unique strengths.

Related work

Understanding human-human and human-AI collaboration

Researchers in CSCL and beyond learning research have begun exploring a variety of approaches comparing human-human collaboration (HHC) with human-AI collaboration (HAI). Tong et al. (2025) found that while both configurations improved physics problem-solving, HHC showed larger effect sizes than HAI, suggesting that peers offer benefits beyond correct answers. In a meta-analysis, Vaccaro et al. (2024) investigated patterns across different human-AI configurations, finding that human-AI teams underperform on decision tasks but excel at content creation, highlighting the nuances in shaping collaborative effectiveness.

Within learning contexts specifically, research examines the experiences of AI-integrated experiences across dimensions. Järvelä et al. (2025) found differences in perceived agency and ownership between peer and AI co-authors. Hernández-Leo et al. (2025) explored how generative AI can improve some types of collaborative workflows while interfering with others. Kosmyna et al. (2025) explored neural and behavioral impacts of AI-supported writing, emphasizing the tradeoff between immediate improvement and long-term learning implications. These comparisons establish precedent for examining HHC and HAI approaches but focus primarily on outcomes rather than interactional processes.

Interaction analysis and real-time mediation in CSCL

Productive collaboration has distinct interactional patterns. Balanced participation is associated with deeper learning and better problem-solving (Barron, 2003). Transactive discourse (building on or challenging others' ideas) promotes more complex understanding than parallel thinking (Berkowitz & Gibbs, 1983; Teasley, 1997). Question-asking drives exploratory talk and collective knowledge construction (Mercer & Littleton, 2007). Constructive controversy, social question-asking, and supported uncertainty promote increased levels of curiosity (Sinha et al., 2017).

The broader field of conversation analytics has developed a wealth of approaches for analyzing collaborative discourse, including using recent LLM advances as an analytic tool. CPSCoach 2.0 (D'Mello et al., 2024) demonstrates real-time classification of collaboration quality using multimodal features, enabling adaptive support during group work. The Jigsaw Interactive Agent (Doherty et al., 2025) uses LLMs to scaffold specific group dynamics like respectful disagreement and evidence-based reasoning. Lee et al. (2025) showed that AI-facilitated small-group collaboration through real-time analytics improved both math learning outcomes and reduced learning anxiety. These systems show the technical feasibility of discourse-aware support but are typically evaluated as interventions rather than instruments for understanding baseline differences across learning ecologies. Our research applies related computational techniques comparatively, aiming to understand how collaborative behaviors like curiosity expressions, responsive questioning, and participatory balance manifest differently when one's partner is an algorithm rather than a peer, and how this impacts both learning and social perception outcomes.

Methods

Participants and learning task

We recruited thirty-six undergraduate students (ages 18-22; 20 female, 12 male, 4 non-binary) from several US universities. The study was approved by institutional IRB. Participants were recruited through university mailing lists and a behavioral research platform, receiving $25 for participation. Students were randomly assigned to either peer dyads (12 pairs, n=24) or individual AI-assisted learning (n=12). Participants in peer dyads were not acquainted prior to the session, mimicking the format of drop-in study support rather than established classroom relationships.

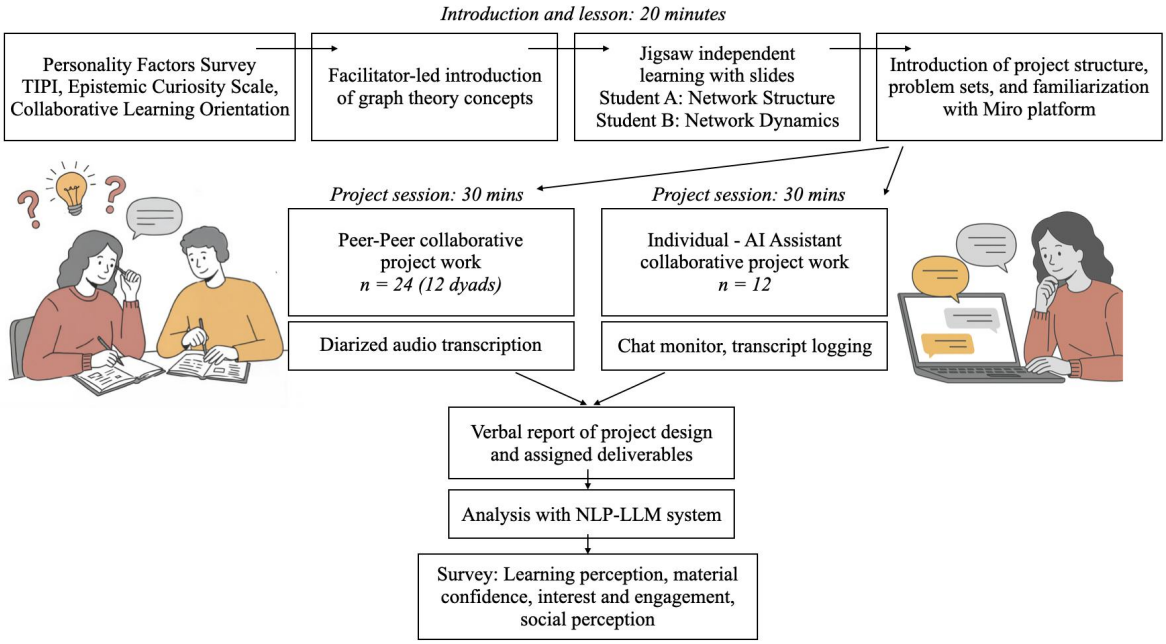
All sessions were conducted remotely via Zoom to reflect increasingly common online learning environments and to constrain interaction channels for discourse analysis. The learning activity centered on graph

theory concepts applied to network design problems, selected for its balance of conceptual understanding and creative problem-solving. This domain requires both knowledge acquisition (understanding graph properties) and application (designing networks), allowing us to observe different interaction dynamics.

Following a brief instructor-led introduction covering basic concepts (vertices, edges, directionality, weighting), participants engaged in individual study. In the peer condition, pairs received complementary learning materials in a jigsaw design: one student studied network structure while the other studied network dynamics. In the AI condition, students received one of these material sets. This design supported comparable knowledge preparation across conditions while creating interdependence in peer pairs.

Students then engaged in a 30-minute collaborative project session designing transportation or communication networks while optimizing for specific constraints (e.g., minimizing connections while maintaining reachability, designing resilience to failure cases, identifying critical nodes). All participants used Miro, an online whiteboard platform, for the design task, after a brief introductory exercise to build comfort with basic drawing tasks on the platform. AI condition participants interacted with a custom chat assistant through a web interface while completing the same design task. At the end of the project phase, students in both conditions were asked to explain their network design and report how they met the task deliverables.

Figure 1
Study Design Overview



Chat assistant design

The AI assistant was built using the Anthropic API (Claude 3.5 Sonnet) with custom prompting designed to support collaborative learning. The interface presented a conversational chat-bubble interaction accessed through a web browser, with chat data transmitted to a secure server for real-time monitoring. The assistant's prompt design built on established educational chatbot frameworks (Mollick & Mollick, 2023), reflecting a likely real-world use case of existing AI chat interactions. The prompt emphasized qualities including:

- Transparency about being an AI assistant
- Support for learning through problem-solving guidance rather than direct answers
- Encouragement of student reasoning and problem-solving; curiosity about student thinking
- Offering structured guidance with clarifying questions

In practice, assistant turns predominantly combined questioning and curiosity probing (asking students for their reasoning or thoughts on next steps) with targeted explanation, consistent with the Socratic guidance framing of the prompt design. The assistant was provided with knowledge of the lesson materials and individual study content, scoping its support accordingly. Chat transcripts were logged with timestamps for subsequent analysis.

Data collection and analysis

 © ISLS

Pre-session measures

Participants completed baseline assessments including the Ten Item Personality Index (TIPI; Gosling et al., 2003), Epistemic Curiosity Scale (Litman et al., 2003), prior experience with online learning and AI tools, and collaborative learning orientation items to test whether individual differences predicted collaboration outcomes.

Session recording

All sessions were recorded via screen capture and audio recording. Peer sessions captured both students' verbal exchanges and their shared Miro workspace. AI sessions captured timestamped chat transcripts, Miro workspace activities, and audio/video recording of verbal commentary including the closing presentation.

Post-session measures

Immediately following the project work, students verbally presented their network designs to the experimenter and completed a 32-item perception survey assessing four dimensions:

- Confidence in learning and material understanding (6 items, *e.g. I feel that I could teach this material to someone else"*)
- Interest in material (5 items, *e.g. I might apply these concepts to other projects"*)
- New ideas and curiosity (5 items, *e.g. I wanted to dive deeper into the material beyond what was taught"*)
- Social perception of learning experience (6 items, *e.g. Working with my partner/AI helped me learn better than I would have alone"*)

Discourse analysis

To enable detailed analysis of interaction patterns across conditions, we developed and validated a conversation analytics system that classifies curiosity expressions (specific, diversive, epistemic, social) and interaction features (questioning, explanation, agreement/disagreement, uncertainty) in learning discourse. The system uses a hybrid architecture in which local NLP classifiers handle high-confidence cases (≥0.6 confidence), while low-confidence segments are escalated to Claude 3.5 Sonnet (Anthropic API) for classification of context-dependent expressions; semantic pattern descriptions are available in Morris & Maes (2025). The classification system was validated against 450 human-annotated conversation segments, with each segment receiving approximately 12 independent human ratings. Inter-rater reliability showed moderate to high agreement (κ=0.72 for curiosity presence, κ=0.76 for interaction type classification, κ=0.66 for curiosity subtype classification). From classified segments, we computed *curiosity expression rates* (normalized instances per conversation), *social curiosity ratio* (social curiosity / total curiosity expressions), *explanation frequency* (explanations given per participant), and *turn-taking balance*.

Statistical analysis

Our analysis examined three levels of relationships: (1) between-group comparisons of survey measures, (2) within-conversation relationships among discourse features, and (3) correlations between conversation-level discourse metrics and post-session survey responses.

For between-group comparisons of survey responses, we used independent samples t-tests with Cohen's d effect sizes, treating 5-point Likert scales as interval data to enable effect size estimations. Groups were formed using median splits within each condition (peer vs. AI) based on collaboration quality perceptions.

To examine relationships between discourse features and learning outcomes, we calculated Pearson correlations at the individual participant level. For peer sessions, discourse features were coded separately for each participant within the dyad (e.g., number of curiosity expressions by participant A, number of explanations given by participant A), with n=24 individual data points. These individual-level discourse metrics were then correlated with each participant's own survey responses. For AI sessions (n=12 individuals), each participant provided one data point.

Given power limitations due to our sample size (n=24 for peer participants, n=12 for AI participants), we report effect sizes alongside significance values with α = .05 for statistical significance. We note that individuals within peer dyads may show correlated outcomes due to their shared interaction experience; however, the exploratory nature of this work and the consistency of patterns across multiple measures suggest meaningful relationships worth further investigation. All analyses were conducted using Python (version 3.10) with scipy.stats and pandas libraries.

Results

Collaboration quality divide

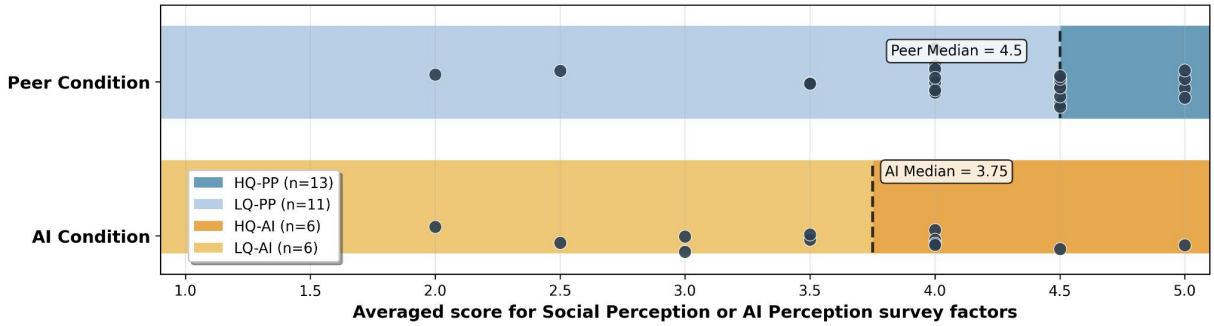
Analysis revealed substantial variation in collaboration quality both between and within conditions. To explore whether this variation corresponded to meaningful differences in discourse patterns, participants were grouped by median split into higher and lower experience groups within each condition (see Figure 2). This approach was motivated by evident distributional clustering; we treat these as descriptive groupings for pattern exploration rather than validated categories. For peer-peer interactions, this distinction emerged from surveyed social perception measures (e.g., quality of collaborative experience, partner engagement). For AI interactions, groupings were based on AI engagement perception (e.g., quality of interaction with the assistant). These groupings revealed a pattern we term the "collaboration quality divide."

High-quality peer-peer interactions *(HQ-PP, n=13)* reported the strongest collaborative experiences (M=4.85, SD=0.38), significantly outperforming both low-quality peer-peer collaborations *(LQ-PP, n=11)* (M=3.82, SD=1.25; t(22)=2.58, p=.017, d=1.16) and all AI conditions (M=3.10, SD=1.20; t(23)=4.53, p<.001, d=2.10). This gap between high-quality peer and AI interactions is consistent with qualitatively different learning experiences rather than incremental differences, though the small sample warrants cautious interpretation.

Low-quality peer-peer interactions performed worse than AI participants on multiple measures, challenging assumptions about the inherent value of peer learning. This pattern held across learning confidence (LQ-PP: M=3.18, SD=0.75 vs. AI: M=4.30, SD=0.67; t(21)=-3.62, p=.002, d=-1.56) and interest in the material (LQ-PP: M=3.36, SD=0.92 vs. AI: M=3.75, SD=0.75; t(21)=-1.08, p=.29, d=-0.46). This divide aligns with ZPD's prediction that scaffolding quality shapes developmental outcomes; peer interactions, unlike AI, offer scaffolding that varies with the genuine uncertainty of the scaffolder.

Figure 2

Participant Ratings of Peer-Learning Perception or AI-Learning Perception



Participants are separated at each condition's median into high quality (HQ) and low quality (LQ) interaction groups per condition.

Peer-peer or individual-AI differential effects

Groups showed distinct patterns across different aspects of the learning experience (see Figure 3). For confidence in explaining concepts ("I feel capable of helping others understand the topic"), a surprising pattern emerged: AI participants (M=4.30, SD=0.67) and LQ-PP participants (M=3.18, SD=0.75) reported the highest and lowest confidence respectively (t(21)=-3.62, p=.002, d=-1.56). This suggests that an AI-based interaction can build subject-material confidence even when deeper engagement is limited, perhaps because of the consistency and factual nature of the AI support.

Conversely, for generating new ideas and curiosity ("I found myself asking new questions beyond the lesson material", "I wanted to explore new ideas", "The project sparked deeper interest"), HQ-PP participants (M=4.00, SD=0.91) significantly outperformed AI participants (M=3.10, SD=0.99; t(23)=2.16, p=.043, d=1.00). The LQ-PP exchanges fell between these extremes (M=3.36, SD=0.92), suggesting that even lower-quality peer interactions may generate more curiosity than AI exchanges. This "confidence without curiosity" pattern in AI interactions reflects their more transactional nature: students get answers efficiently but miss the generative friction of peer collaboration.

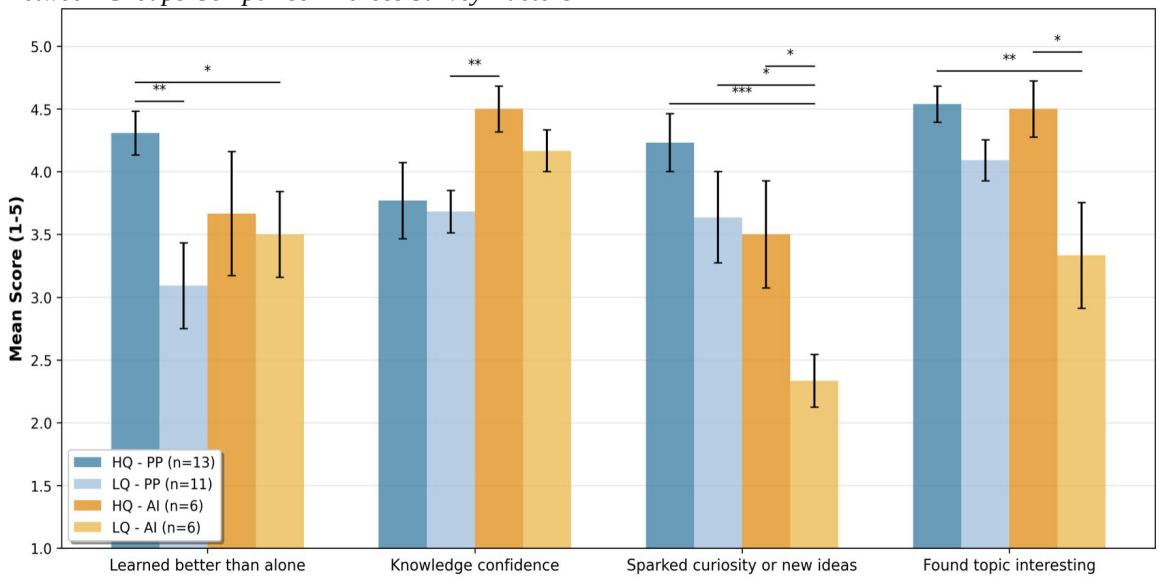
Discourse patterns as predictors

Discourse analysis on peer pairs reveals relationships between specific interaction patterns and learning perceptions (Table 1). While our sample size limits statistical power (n=24 for peer condition analyses), the observed correlations suggest meaningful relationships warranting further investigation, and show consistency with the theoretical role of social curiosity as a driver of generative questioning (Sinha et al., 2017).

 © ISLS

Figure 3

Between-Groups Comparison Across Survey Factors



Comparison of four groups across survey measures from categories (L-R): peer/AI learning value, knowledge confidence, curiosity, and interest. From L-R in each category: HQ-PP group, LQ-PP group, HQ-AI group, LQ-AI group. *Note*: *p<.05, **p<.01, ***p<.001

Table 1

Emergent Interaction Features and Correlations to Curiosity, Interest, and Survey Factors.

Interaction feature	Curiosity / Perception component	r	p
Question asking	Overall curiosity rate	0.565	0.0041**
Question asking	Social curiosity rate	0.642	0.0007***
Explanation giving	Social curiosity rate	0.439	0.035*
Social curiosity rate	Asked new questions (Q26)	0.468	0.022*
Social curiosity rate	Understanding of concepts (Q21)	0.439	0.035*
Explanation giving	Value of collaboration (Q17)	0.417	0.039*
Overall curiosity rate	Understanding of concepts (Q21)	0.381	0.062

*Note: *p<.05, **p<.01, ***p<.001*. Only the most relevant correlations are shown.

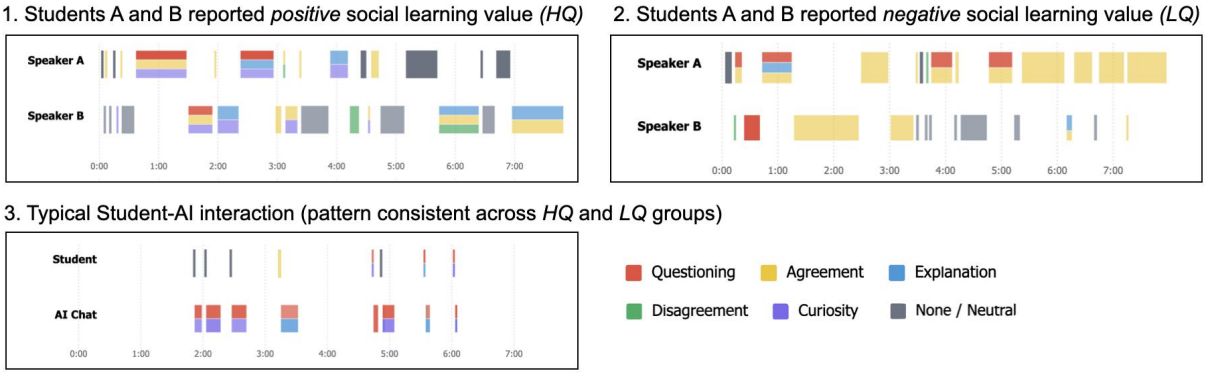
Social curiosity expressions—interest in the partner's perspective or reasoning—showed notable correlations with learning outcomes. Students who expressed more social curiosity reported asking new questions beyond the lesson material (r=0.468, p=.022) and better understanding of graph theory concepts (r=0.439, p=.035). Additionally, explanation-giving behavior positively correlated with perceived value of collaboration (r=0.417, p=.039). The strong correlation between question-asking and social curiosity (r=0.642) likely reflects overlapping constructs in our coding, as social curiosity was frequently enacted through questions about partners' perspectives (e.g., “Why did you do it that way?”), an expected behavior in peer dialogue.

Qualitative analysis of discourse revealed distinct patterns distinguishing high- from low-quality collaborations. Sessions receiving high social perception ratings displayed specific discourse markers early in interactions: frequent questions about partner reasoning (“Why did you connect those nodes that way?”), explicit building on partner contributions (“Like you mentioned about redundancy, but what if we also considered…”), and verbal acknowledgment of learning (“Oh, I hadn't thought about that”).

Figure 4 illustrates these contrasting patterns. High-quality conversations showed balanced turn-taking, diverse interaction types (questioning, explaining, disagreement), and greater presence of social curiosity expressions. Low-quality conversations, despite sometimes having balanced turn lengths, consisted primarily of agreement or neutral statements without deeper engagement; participants were often agreeable but not genuinely collaborative.

Figure 4

Comparison Between HQ-PP, LQ-PP, and Student-AI Conversations



Discourse patterns in representative conversations. Horizontal bars show speaker turns over time (width = turn length) with two rows per conversation representing each speaker. Colors indicate coded interaction features (see legend). Vertically stacked colors within a single bar indicate multiple features in one turn; for example, "Yeah, I think you're right. Look, we can also do that with an undirected graph like this, though. How should the shuttle work here?" would be identified as containing *agreement*, *explaining*, and *questioning*. AI chat turn length reflects reading time derived from message length.

Turn-taking and interaction asymmetry

We also noted differences in turn taking, both between positive and negative perception groups per condition, as well as between modalities. HQ-PP dyads demonstrated more balanced contributions (turn ratio ≈ 0.45-0.55), while LQ-PP students showed greater imbalance (turn ratios ranging 0.20-0.80). AI sessions showed extreme asymmetry: students provided minimal responses (M=5 words per turn) compared to AI's lengthier explanations (M=27 words).

Qualitatively, the nature of these interactions differed as well. Students in AI conditions often ignored or resisted collaborative prompts from the assistant, treating it as an information source rather than a collaborative partner – a pattern visible qualitatively across sessions. One participant explicitly stated: "stop asking me questions, I just want to know what you think." Throughout AI sessions, we observed students providing brief confirmations ("okay," "got it") rather than engaging with the assistant's attempts at dialogue, questioning, and curiosity about the student's approach.

These patterns were also clear in the final presentations: a majority of AI participants (9 out of 12) used exact terminology from assistant responses in their presentations, also using more technical language ("This uses adaptive routing for high-capacity paths"), mirroring the AI assistant. Several students, when asked for details about their decision making, acknowledged their more passive role: "The chat suggested that I include that, and it seemed like a good idea".

In contrast, peer participants demonstrated more ownership of their solutions. Their presentations emphasized the collaborative process: "So, first we thought about having a central hub, but then we realized that we should use a triangulated system…" Even during presentations, they negotiated who would explain what: "Do you want to describe how we did the weighting?" Peer participants used less technical jargon, instead explaining concepts in their own words.

Interaction quality is emergent, not individual

Notably, interaction quality did not appear to be predicted by individual traits. Personality measures collected pre-session (TIPI dimensions, curiosity orientation, collaboration preference, topic interest) showed no significant correlations with any outcome variables (all $p>.20$). This held true across both conditions, suggesting that collaboration success depends on what unfolds between partners, not who they are individually.

The AI condition provided additional evidence for this trend. Despite receiving nearly identical chatbot behavior, participants diverged into the HQ-AI group (n=6) and LQ-AI group (n=6) based on their interaction

patterns. The high-engagement group maintained higher question frequency and attempted to build on AI suggestions, while the low-engagement participants adopted mostly transactional patterns, ignoring questions and prompts from the AI assistant. These behavioral differences corresponded to perception outcomes comparable in magnitude to the peer-condition divide, suggesting engagement patterns matter regardless of modality.

Discussion

Complementary affordances and the nature of collaboration

Our initial findings suggest that the value of peer learning depends critically on interaction quality, not inherent superiority of a single mode. The collaboration quality divide demonstrates peer learning as high-risk, high-reward: when effective, it significantly outperforms AI on engagement and curiosity metrics, but when it fails, students might be better served by AI assistance. AI excels at consistent knowledge delivery and building procedural confidence; students felt capable of explaining concepts despite showing less engagement. This "confidence without curiosity" pattern reflects transactional human-AI interaction where students receive answers efficiently but miss the generative friction of navigating uncertainty together. Notably, students naturally adopted these different interaction modes. They didn't treat AI as a collaborator despite its attempts at Socratic dialogue; they used it as a tool, suggesting systems should embrace rather than mask their tool nature.

Theoretical implications

These initial findings point toward extending Vygotsky's ZPD concept into AI-mediated contexts. While AI provides scaffolding, it may not activate the same socio-cognitive mechanisms as peer interaction. The strong correlation between social curiosity and learning outcomes suggests that curiosity about others' thinking processes may be as important as content curiosity itself, a dimension that AI interactions appear less likely to activate.

These patterns also illuminate why the frameworks introduced earlier matter empirically. The extreme turn asymmetry in AI conditions suggests socially shared regulation was structurally absent: rather than co-monitoring understanding and jointly adjusting approaches, students received information and confirmed receipt. The corresponding absence of transactive discourse – extensions and challenges replaced by brief confirmations – likely explains the confidence-without-curiosity pattern. Terminological knowledge can be passively encoded from consistent, authoritative explanation; the conceptual depth that transactive exchanges generate requires the reciprocal knowledge-building that was absent here. Together, these suggest the collaboration quality divide reflects not just subjective experience but the differential activation of the socio-cognitive mechanisms Vygotsky, Järvelä and Hadwin, and Teasley identify as central to deep learning: scaffolding through genuine uncertainty, co-regulation of understanding, and cumulative knowledge-building through transactive exchange.

Design implications and limitations

The discourse markers that suggest interaction quality – asking questions, showing social interest, giving explanations – appear early enough for real-time intervention. Rather than AI replacing peers, it could monitor collaborations and provide support when struggling patterns emerge (extreme turn imbalance, low questioning rates). Future systems might combine modalities: leveraging peer interaction for curiosity generation and AI for knowledge consolidation when peer dynamics falter.

We acknowledge several limitations within our research. Sessions were single encounters lasting approximately one hour; longer-term collaborations might show different patterns, particularly as relationships develop and students learn to work together. The specific graph theory task and remote format may not generalize to other learning contexts or face-to-face settings. Additionally, the jigsaw design created structural knowledge interdependence in peer pairs that was absent in the AI condition; future work should test conditions with matched material access to isolate modality effects. Additionally, the discourse analysis system's moderate inter-rater reliability reflects both technical limitations and inherent subjectivity in identifying curiosity expressions. Our sample sizes for AI subgroups were small, though they showed initial consistent patterns across multiple analyses.

Conclusion

As educators integrate AI into classrooms, the question isn't whether it should replace peer learning but how it can strengthen it. By identifying early markers of collaboration quality and understanding the complementary affordances of peers (sparking ideas, generating interest and curiosity) and AI (consistent knowledge delivery), we can design environments that preserve social learning benefits while leveraging AI's strengths. Our findings suggest that effective inclusion of AI in education goes beyond seeing human or AI collaborators as uniformly beneficial or detrimental, but rather detecting and nurturing quality interactions wherever they emerge.

References

Barron, B. (2003). When smart groups fail. Journal of the Learning Sciences, 12(3), 307–359. https://doi.org/10.1207/s15327809jls1203_1

Berkowitz, M. W., & Gibbs, J. C. (1983). Measuring the developmental features of moral discussion. Merrill-Palmer Quarterly, 29(4), 399–410.

Cacioppo, J. T., & Cacioppo, S. (2018). The growing problem of loneliness. Lancet, 391(10119), 426. https://doi.org/10.1016/S0140-6736(18)30142-9

Chi, M. T. H., Siler, S. A., Jeong, H., Yamauchi, T., & Hausmann, R. G. (2001). Learning from human tutoring. *Cognitive Science*, *25*(4), 471–533. https://doi.org/10.1207/s15516709cog2504_1

D'Mello, S. K., Duran, N., Michaels, A., & Stewart, A. E. B. (2024). Improving collaborative problem-solving skills via automated feedback and scaffolding: a quasi-experimental study with CPSCoach 2.0. User Modeling and User-Adapted Interaction, 34(4), 1087–1125. https://doi.org/10.1007/s11257-023-09387-6

Doherty, E., Perkoff, E. M., von Bayern, S., Zhang, R., Dey, I., Bodzianowski, M., Puntambekar, S., & Hirshfield, L. (2025). Piecing together teamwork: A responsible approach to an LLM-based educational jigsaw agent. Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems, 1–17. https://doi.org/10.1145/3706598.3713349

Gosling, S. D., Rentfrow, P. J., & Swann, W. B., Jr. (2003). A very brief measure of the Big-Five personality domains. Journal of Research in Personality, 37(6), 504-528. doi: 10.1016/S0092-6566(03)00046-1.

Hernández-Leo, D. (2025). Collaborating with generative AI for learning. In International Society of the Learning Sciences (pp. 525–533).

Järvelä, S. (2025). Exploring human-AI interactions across various contexts. In International Society of the Learning Sciences (pp. 481–489).

Järvelä, Sanna, & Hadwin, A. F. (2013). New frontiers: Regulating learning in CSCL. Educational Psychologist, 48(1), 25–39. https://doi.org/10.1080/00461520.2012.748006

Kosmyna, N., Hauptmann, E., Yuan, Y. T., Situ, J., Liao, X.-H., Beresnitzky, A. V., Braunstein, I., & Maes, P. (2025). Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for essay writing task. In *arXiv [cs.AI]*. arXiv. http://arxiv.org/abs/2506.08872

Lee, M., Feng, M., & Miller, M. (2025). Collaborative math learning facilitated by an intelligent agent. In Proceedings of the International Society of the Learning Sciences (pp. 1-10).

Litman, J. A., & Spielberger, C. D. (2003). Measuring epistemic curiosity and its diversive and specific components. *Journal of Personality Assessment*, *80*(1), 75–86. https://doi.org/10.1207/S15327752JPA8001_16

Malmberg, J., Järvelä, S., & Kirschner, P. A. (2015). Elementary school students 'strategic learning: Does task-type matter? Metacognition and Learning. 10, 113–136.

Mercer, N., & Littleton, K. (2007). Dialogue and the development of children's thinking: A sociocultural approach. Routledge. https://doi.org/10.4324/9780203946657

Mollick, E. and Mollick, L. Assigning AI: Seven Approaches for Students, with Prompts (September 23, 2023). The Wharton School Research Paper, Available at SSRN: https://ssrn.com/abstract=4475995 or http://dx.doi.org/10.2139/ssrn.4475995

Morris, C. and Maes, P. (2025). MoSAIC: A Hybrid Framework for Understanding Social Curiosity in Digital Learning Environments. figshare. Poster. https://doi.org/10.6084/m9.figshare.31871980.v1

Pentina, I., Hancock, T., & Xie, T. (2023). Exploring relationship development with social chatbots: A mixed-method study of replika. Computers in Human Behavior, 140(107600), 107600. https://doi.org/10.1016/j.chb.2022.107600

Roscoe, R. D., & Chi, M. T. H. (2007). Understanding tutor learning: Knowledge-building and knowledge-telling in peer tutors 'explanations and questions. Review of Educational Research, 77(4), 534–574. https://doi.org/10.3102/0034654307309920

Sinha, T., Bai, Z., & Cassell, J. (2017). Curious minds wonder alike: Studying multimodal behavioral dynamics to design social scaffolding of curiosity. In arXiv [cs.HC]. arXiv. http://arxiv.org/abs/1705.00204

Teasley, S. D. (1997). Talking about reasoning: How important is the peer in peer collaboration? In Discourse, Tools and Reasoning (pp. 361–384). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-03362-3_16

Tong, D., Zhang, J., & Chen, X. (2025). Exploring the role of human-AI collaboration in solving scientific problems: Comparing human-AI with human-human collaboration. Physical Review Physics Education Research, 21(1).

Twenge, J. M., Haidt, J., Blake, A. B., McAllister, C., Lemon, H., & Le Roy, A. (2021). Worldwide increases in adolescent loneliness. Journal of Adolescence, 93(1), 257–269. https://doi.org/10.1016/j.adolescence.2021.06.006

Vaccaro, M. A., Almaatouq, A., & Malone, T. (2024). When are combinations of humans and AI useful? In *arXiv [cs.HC]*. arXiv. https://doi.org/10.31234/osf.io/mk46p

Vygotsky, L. S. (1978). Mind in Society: The Development of Higher Psychological Processes. Cambridge, MA: Harvard University Press.

Acknowledgments

This research was supported by a fellowship from the MIT Morningside Academy for Design (MAD) and a grant from Cosmos Institute. Participant management was supported by the MIT Behavioral Research Lab.

International Society of the Learning Sciences

Frame Alignment and Feedback in CT-ified Science Classrooms

Connor Bain, Uri Wilensky
connor.bain@northwestern.edu, uri@northwestern.edu
Northwestern University

Abstract: While computational thinking (CT) is an integral part of modern scientific inquiry, STEM teachers struggle to incorporate these ideas in classrooms. We report on the efforts of a design-based implementation research project in which teachers participate in a constructionist co-design process to design and then implement CT-ified STEM units in their classrooms. Using feedback teachers left on student responses from these unit implementations, we utilize a basic bag-of-words model to show that teachers did specifically attend to student learning at the intersection of CT and STEM. We then further explicate this evidence through a qualitative think-aloud interview with one of the teachers who designed and implemented a curriculum based around computational models in biology. What we find is evidence that STEM teachers in CT-ified classrooms can assess both CT and STEM skills at the same time though this intersection still needs additional attention in terms of tools and PD for teachers.

Introduction

> "There was a time when I believed, as many people do, that teachers would be the most difficult obstacle in the way of transforming School…this simplistic belief is, in reality, a far greater obstacle to education change than the fact some teachers are actually conservative." (Papert, 1993)

Computational Thinking (CT) practices have historically been featured mostly in standalone computer science (CS) courses, resulting in unequal access for students from historically underrepresented groups in CS, such as women and racial minorities (Margolis & Fisher, 2003). CT has become a necessary and integral part of nearly every discipline, particularly STEM disciplines (Weintrop et al., 2015) given the near ubiquity of computation. The Next Generation Science Standards have made clear that using computational thinking (CT) is a cornerstone of modern science education (Quinn et al., 2012; Wilensky, Brady & Horn, 2014). By embedding CT practices into high school STEM classrooms like biology, chemistry and physics, we can simultaneously improve access to CT for all students, particularly those underrepresented in CS, while also providing a more authentic STEM experience for students in these classes.

Here we describe work done in a research practice partnership between a Midwestern U.S. research university and several urban high schools in a large Midwestern city. Previous professional development programs focused on "CT" have taken various forms, each choosing to treat CT as either a pedagogical framework or as content. For example, Brennan & Resnick (2012) treat computational thinking concepts as new content (e.g., sequences, loops, parallelism, etc.) to be presented to students through the creation of digital hypermedia projects in the Scratch programming language. On the other hand, Mouza et al. (2017) treat CT as a pedagogical frame by introducing pre-service teachers to the basic concepts of CT and helping them design lessons that incorporated CT concepts. While both works aim to help teachers integrate CT into existing classrooms, they each go about it in a very different manner. The approach we follow here aims to treat CT as both a set of new concepts as well as a pedagogical frame that allows teachers to integrate CT throughout their curriculum (Weintrop et al., 2015).

Though other work has focused on evaluating student learning and PD design (Kelter et al., 2021; Peel et al., 2020; Wu et al., 2020), here we are more focused on the teacher's perspective of the alignment of CT practices with their classroom's underlying STEM learning objectives–specifically when it comes their own evaluations of student outcomes (Bain et al., 2020). Specifically, we ask the question, how do science teachers attend to student learning in both computational thinking and STEM-content knowledge in a "CT-ified" science classroom? By analyzing data quantitative and qualitative methods from the 2nd iteration of a design-based implementation research (DBIR) project, we show that when tasked with integrating CT ideas into existing STEM coursework, teachers attend to both CT learning and STEM-content learning in their classrooms. The implications of this finding show that, given structured professional development opportunities, teachers see assessing student learning in both STEM and CT practices as part of their pedagogy.

Background

 © ISLS

Computational thinking professional development

This work is part of a larger DBIR project that has already undergone two full year cycles. Within each cycle, first the research team integrated the three overarching goals: 1. Engage students in authentic science practice; 2. Deepen scientific content understanding; 3. Broaden participation in computing, into a 4-week professional development program. These goals are framed as student-level outcomes, yet the project's approach to intervention at the teacher-level means that teachers are framed as the main changemakers in each classroom. This is a purposeful choice which shows a clear philosophical stance when it comes to the role of the teacher in education: *that content/curricula alone are not enough to drive deep scientific content learning*. We see the teacher as Papert came to see them later in his research–as a world builder, learning guide, and changemaker. As such, it is difficult to imagine any student outcomes without considering how teachers might enable those outcomes in their classroom.

The project found that very little work has been done to study helping teachers in CT-integration (Sands et al., 2018) and gradually pivoted to designing PD that focused on constructing CT-ified STEM units in a co-design setting (Kelter et al., 2021; Peel et al., 2020; Roschelle et al., 2006), where teachers, framed as pedagogical and content experts, and researchers, framed as computational and educational design experts, worked together to build CT-ified units and explore what CT could look like in each teacher's unique classroom. The role of researcher varied from co-design team to co-design team, with some researchers taking on a role more akin to a computational expert. However, it is important to note that this co-design format is not the same as previous software focused co-design projects, e.g., (Roschelle et al., 2006). The researchers were not software engineers and teachers were always encouraged to take part in the development of any computational tools used in their curricula. In this sense, all team members were working together to construct not simply new curricula, but also new understandings of what the modern science classroom could look like.

Teacher practice

To structure our conversation of practice, we borrow from Windschitl et al.'s (2012, p. 882) definition of teacher practice, specifically its mechanism for dividing practice into a three-step cycle: planning, enacting, and reflecting. While Windschitl et al. did not mean this to be an exhaustive taxonomy of practice, we leverage it here as a framing device to breakdown how we might study emergent teacher practice in a new type of classroom. The final part of teacher practice, *reflecting*, we see as grounded in not only self-reflection of the teacher on their own practice (e.g., "how do I think today went?") but also as the process of tracking student learning across an implementation (i.e., looking at and assessing student work for progress toward learning goals). It is this portion of the cycle we focus on here to answer our question: do teachers in such a CT-ified science classroom attend to learning in both CT and STEM content?

Methodology

The main data at play in this work is the feedback teachers provide students within each CT-ified STEM unit. Teachers had the ability to leave individualized feedback for each student on each response to each question in their units. It is important to note that while all teachers were shown how this function could be used both live in their classroom and after students had submitted their work, they were not required to use this function at any point.

Data collection

Fifteen teachers participated across the two cycles of the project representing Biology, Chemistry, Environmental Science, Physics, and Statistics. They represented six different high schools in the greater area surrounding a large Midwestern city in the U.S. From these fifteen participants came twelve newly designed CT-ified STEM curricular units, each roughly two-weeks long.

Students participated in each unit developed by our participating teachers via a web-based learning management system (LMS). Teachers have access to each student response and can provide both point-based and text-based feedback to students' responses. Thus, our main corpus of data is the feedback teacher's chose to leave students. Across the two years of the study, 1219 students participated in CT-ified units contributing 106050 responses to questions. Of these responses, teachers only provided written feedback (i.e. a comment) on only 5654 responses. While the proportion of responses receiving feedback may seem low, the interviews we performed with teachers after their implementation indicate that this is part of their typical practice. This analysis does not capture in-the-moment feedback teachers provided in the classroom. At no point were any participating teachers were required to give feedback via LMS so this dataset must be considered using a sort of "opted-in" mindset. It is important to note that this does *not* indicate that teachers were not probing students' understandings–just that

 © ISLS

they chose not to do so via LMS. For example, one of our instructors, Joshua, did not use the feedback mechanism on the website, yet in his post-interview, he specifically mentioned assessing students' understanding:

> …any little bit is going to get in the way of them experiencing the experience they're supposed to have but, like, for the most part of me I'm telling you dude I can't believe how smoothly things went and how frequently when I would check in with students about something they like had the answer I thought they should have, and you know the answers that I looked at, you know as we looked through their work.

Targeted feedback

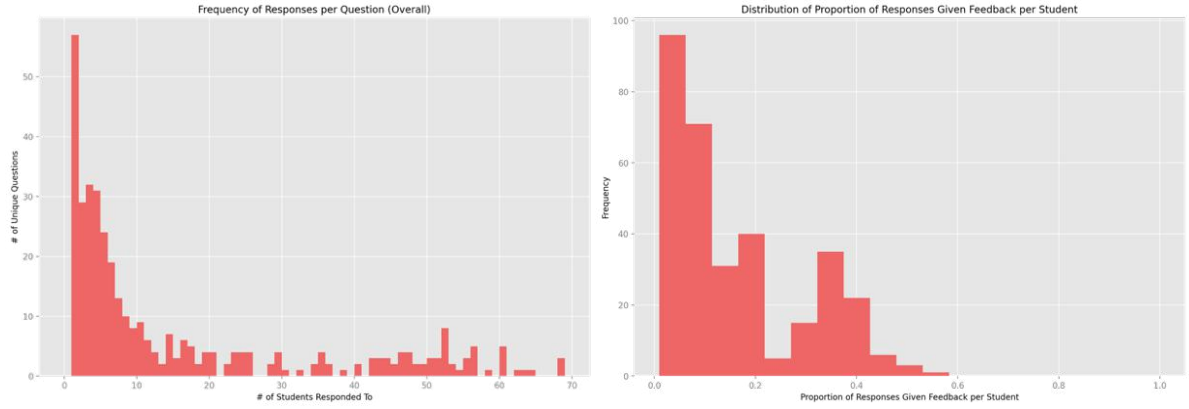
Given that roughly 5% of responses received comment-style feedback, a reasonable question to ask is *which questions receive feedback*? Recall our research question: *how do science teachers attend to student learning in **both** computational thinking and STEM-content knowledge in a "CT-ified" science classroom*? In this section, we analyze the content of questions where teachers decided to leave comments on student responses. By using some techniques from natural language processing and correlating them with qualitative post-implementation interview responses, we begin to build a picture of how and why teachers target their feedback in these CT-ified classrooms.

Question selection

For example, when teachers do decide to give feedback on a question, do they typically give all students feedback for one target question? In Figure 1, we visualize the distribution of number of students responded to across unique questions. In this figure, the height of the bar represents the number of unique questions that teachers gave feedback to according to the number of students on the x-axis. For example, the left-most bin with a height of 56 indicates that there are 56 unique questions across our data where only a single student received feedback. The right-most bin on the other hand says that there are 4 unique questions where 68 students received feedback.

Figure 1

Distribution of the Number of Feedback Entries Per Question (a); Feedback Given Per Student (b).



(a) (b)

The distribution is quite skewed to the left with a secondary mode centered around 50 students. What this shows is that teachers seem to be engaging in two separate strategies: 1. Spot checks where either particular students are targeted for feedback or responses to some particular question are scanned but not everyone receives feedback; 2. Specific focus questions where teachers give feedback to all students for those particular questions. In Figure 2, we look at a related statistic: the distribution of the proportion of responses for which each student received feedback. Even after omitting the students who received no feedback at all, we still see a skewed distribution. Though there is also a secondary mode at the 35% (0.35) mark. This would seem to reinforce the idea that teachers are very targeted with their feedback.

Questions that are most targeted

With this in mind, we now turn to the content of questions that did receive feedback. By using a bag-of-words model (Bain et al., 2021; Wallach, 2006), we attempt to identify similarities in questions targeted by teachers for

 © ISLS

feedback. In Figure 3, we look at the most common words (removing common English stop words like articles) among questions that did receive feedback. Here we see that a number of CT relevant terms are particularly well represented in these targeted questions including "model" (rank 1), "simulation" (rank 4), "data" (rank 14), and "code" (rank 15). We also see several scientific practice terms: "think" (rank 2), "explain" (rank 10), and "relationship" (rank 20). This would seem to indicate that teachers are indeed targeting questions for feedback that address CT skills as well as science content. In fact, if we compare these most common words to the most common words across all questions in all of the units (regardless of whether or not they were targeted for feedback), we see the CT-related terms are overrepresented in the targeted set with "model" remaining at rank 1 and "data" at rank 2. An anlysis at the bigrams of those same questions show "netlogo code", "run simulation", "collect data", "run model," and "computational tools" as the most frequently ocurring bigrams in these questions. This is another indicator that our teachers are indeed attending to both computational thinking and STEM content in the classroom through their feedback practice.

However, the trends in questions targeted for feedback as well as the quantity of feedback itself varies greatly across individual teachers. This could just correspond to existing pedagogical practice. In other words, by studying feedback across a single teacher, we might gain additional insight into teachers' process of reflection. In the next section, we focus in on the teacher who gave the most feedback across the classrooms: Edna.

Focal teacher – Edna

Edna was in her 3rd year teaching biology at a suburban high school in the Midwestern U.S. During the PD, Edna worked on a two-week curriculum centered on the central dogma of molecular biology: that the flow of information is from DNA (information storage) to RNA (information transfer) to protein (functional product). Edna was particularly targeting two NGSS standards, HS-LS1.1 (construct an explanation based on evidence for how the structure of DNA determines the structure of proteins which carry out the essential functions of life through systems of specialized cells) and HS-LS3.1 (ask questions to clarify relationships about the role of DNA and chromosomes in coding the instructions for characteristic traits passed from parents to offspring) (NGSS Lead States, 2013).

Throughout the curriculum which spanned eleven, 42-minute classes, students used microscopes to look at human hair, discussed how genes could express in different ways (and how they are passed on), and also explored computational models of melanin production (Finzer, 2016), DNA/RNA synthesis (using the agent-based modeling language NetLogo), and protein synthesis via a Scratch program (Maloney et al., 2010; Wilensky, 1999).

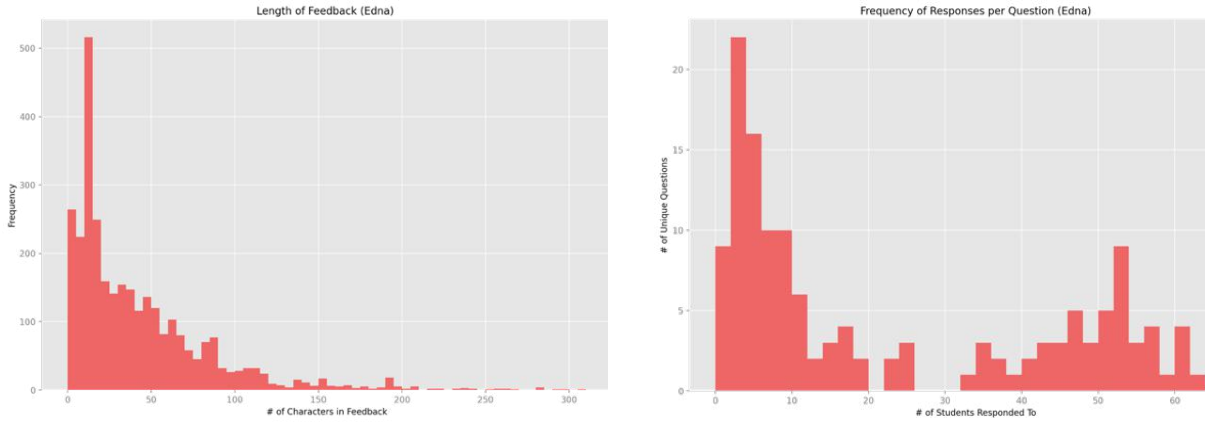
Assessment interview protocol

Rather than just look at Edna's feedback in the same quantitative fashion as our earlier analysis, here we also rely on a think-aloud assessment interview to get a better understanding of how Edna, as a teacher, provides feedback on student work. The interviewer asks the teacher to first walk them through the process of giving feedback on the student work from the CT-ified science unit, occasionally prompting along three lines of inquiry: 1. Question selection; 2. Science content focus; 3. CT content focus. Then the teacher would do the same for a "traditional assignment" (as defined by them) in their course. Finally, the interview ends with a discussion of how this post-implementation feedback differs (if at all) from the just-in-time feedback the teachers provided in their classroom.

Edna's attitude toward assessment is generally directed more at mechanism than memorization when it comes to STEM content learning: "*I don't particularly care about recall of things. I want to know how things are being used, so I don't do a lot of tests. And when I do have tests, you get to bring a "sheet of paper" with anything you want on it.*" Her general strategy was to go through student submissions question by question, giving students feedback: "*What I have been doing is going mostly by question. Okay, and that, so I can be consistent between students. and honestly, sometimes I will copy and paste feedback if I find myself getting the same kind of thing.*" However, we see (*Figure 2b*) the same trend in emulated here that we saw earlier in the overall data set: a few questions with most students getting feedback and a separate group of questions with just a few students getting feedback.

We see the secondary mode at around 50 students (close to Edna's total student count) with a good number of questions targeted for feedback across the entire class. The reason we see the variation around that secondary mode is due to the fact that not all students submitted answers to these particular questions.

Figure 2
(a) Distribution of Feedback Lengths for All Feedback Given by Edna; (b) Distribution of Number of Students Responded to per Question That Received Feedback for Edna.



(a) *(b)*

When we asked Edna to go through her normal assessment process during the interview, she selected a Lesson that had just been completed by students in class on *DNA Transcription and RNA Translation*. She opened the students' responses on the LMS and as she did so, explained:

> Built into this unit, we put in these Summative Assessment Questions [SAQs] these the ones that I've told them that I'm going to look at. What I've been doing is like starting at the that SAQ and then going back up to like where like the first couple of questions here are like review or just brainstorming and I might glance at those later ones.

This accounts for Edna's bimodal distribution of feedback per question. While she uses the SAQs as focal points for all students, she also "glances" at the prior questions to do spot checks as we had surmised. Of course, this begs the question: What do these SAQs look like? In *Table 1*, you can see a number of exemplar SAQs from this particular lesson.

Table 1
Four Exemplar Summative Assessment Questions (SAQs) from Edna's Lesson on Transcription and Translation.

SAQ: What do you notice about the 4th and 5th amino acids and 4th and 5th codons in the sequence you translated? How can you explain what you saw (hint: look back at the codon tables)?
Using the information from this page and the last one, write out the algorithm (the instructions) that you, acting as the ribosome followed to put together the amino acids in the correct order to make a protein based on the mRNA sequence. Be specific!
Think about each of the three models (NetLogo, Concord, Scratch). What is good about each (minimum 1 about each)?
What is the relationship between the ribonucleotides on the mRNA and the amino acids that are being put together to form the protein (think about numbers of things)?

It is this last SAQ that Edna focused on during her assessment interview. As Edna began to go through responses, she narrated her progress:

> It's mostly the process of translation so there's a little bit of review of transcription built in so that's the RNA synthesis that I'm looking for.
> ..
> Okay, so here's what is the relationship between nucleotides and the mRNA and the amino acids that are being put together to form the protein. Think about the numbers of things so with that I was trying to prompt them to think about like the three to one relationship of three nucleotides they don't have the term codon yet but they'll get it….

 © ISLS

She proceeded to quickly go through each students' response looking for evidence of a 3:1 relationship somewhere in each students' reasoning. For a number of students, she found that they described an ordinal relationship (i.e. "there is more of one than the other") and left feedback prompting them to specifically think about the numeric relationship. For some students that she said were "almost there," she would specifically cue the 3:1 relationship in her feedback.

As she finished the first class' worth of student responses, the interviewer prompted her on how she decided to focus on that particular SAQ:

> So we've had these like little check ins [during class] that I've already kind of seeing where they are. [Out of all my questions] probably 50% are application questions and then maybe like 30% just strict biology concept questions and 20% analysis. [The ones that I choose] are…most of them are focusing on biology, a lot of them, though, like, it's more information, I think, for me, then it is like a right answer biologically…

While giving feedback, Edna worked with a number of sources of data. For example, she discussed a number of the questions from this particular lesson with all of the students in the classroom, so she already had a "feel" for how students were thinking about some parts of the topic. In that sense, she targeted this SAQ as a way of focusing on one particular biological facet: the 3 to 1 numerical relationship between the ribonucleotides and amino acids. However, it is not fair to say that Edna is focusing only on biological content in her assessment. Directly after the above, she went on to select another SAQ to focus on "Think about each of the three models (NetLogo, Concord, Scratch). What is good about each (minimum 1 about each)?"

> [As long as they list any good things on all 3] they're good to go. So I'll give them some feedback and am usually looking for some sort of commonality of like what I like about one of them
> …
> yeah for this one [points to a student's response] is just one good thing about each so it would be what I like about some of them….So, like the [other model] gives you a better visual for the process of putting the proteins together like that's really nice, I find that really helpful too, so I might just be like "Oh, I really like that too" for a question like this.

While not an example of a "rigorous" assessment of students' engagement in CT, this question shows a willingness to ask students to reflect on the process of learning with computational models in her biology classroom. In her responses to students, she was careful to both attend to the models' ability to demonstrate/represent the underlying science but also what sort of agency it gave students. For example, when responding to students who commented on the NetLogo model being the most "controllable and organized," Edna left feedback that linked this controllability back to the underlying code of the model: "*I like that about the NetLogo, too (that and I kind of understand how the coding works for them).*" Again, we see, even in feedback, Edna linking the students' scientific experiences back to the underlying computational experiences

If we look at the content of Edna's feedback to students, we also see evidence of feedback pertaining to computational engagement as well as biological understanding. Using Latent Dirichlet Allocation (Blei et al., 2003; Gottipati et al., 2018; Hoffman et al., 2010) with each feedback entry treated as a document, we produced a topic model with five distinct topics, each with fairly high discriminatory power. The five topics that were extracted were most defined by the words shown in *Table 2* using Latent Dirichlet Allocation (with 5 passes and an eta value equal to 1% of the number of topics and alpha values set at auto in the *gensim* Python package (Rehurek & Sojka, 2010)). Across all 5 topics, we see strong evidence that Edna focuses her feedback on the underlying biological content of the curriculum. With the exception of Topics 1 and 2, each of the other three topics is strongly defined by at least one biological concept contained in the unit. Each topic also shows Edna's desire to include affirmation in feedback: "good start," "great," "exactly," "make sense," etc.

Table 2

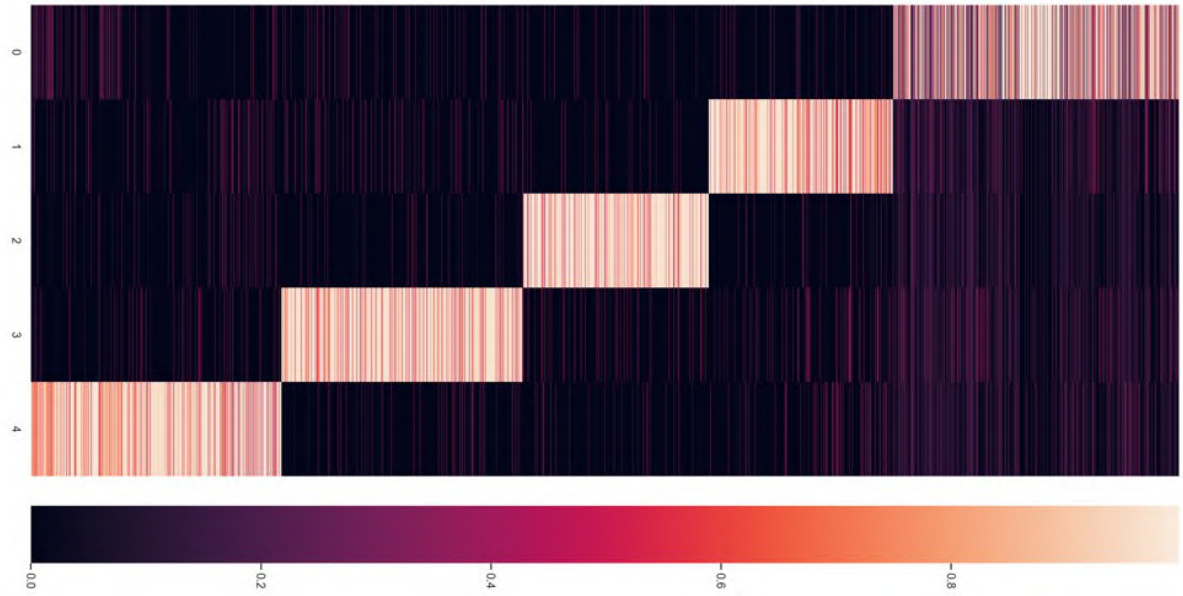
The Five Most Discriminative Words for the Five Topics Extracted via Latent Dirichlet Allocation.

Topic 0	Topic 1	Topic 2	Topic 3	Topic 4
good	great	exactly	nice	amino
start	hair	makes	DNA	acids
C [cytosine]	made	sense	strand	code
G [guanine]	questions	rate	polymerase	definitely

 © ISLS

like	part	interesting	remember	many

Figure 3
A Heatmap Visualizing the Distribution of Topics Across All of Edna's Feedback Entries.



Only Topic 4 includes a word that seem to directly address computational thinking practice (e.g., “code”). In *Figure 3*, the lighter the color, the more associated with the topic on the x-axis the feedback entry is. The entries are sorted based on the topics with which they are most strongly associated. Topics 0 and 4 have the most documents associated with them with Topic 4 being most prolific. If Edna was specifically attending to CT Practices within her feedback, we would expect to see more discriminatory power in CT-relevant words. As we do not see this, it implies that even Edna, who is quite adept at framing students’ engagement with computation, still needs additional knowledge or assistance to directly give feedback on CT as part of her assessment practice. When asked to reflect specifically on her assessment of CT in these units, Edna started off by describing her computational identity:

> I am not a computer person so I think that it's more identity based. It was like before I went to this program in the summer, if you told me that I was going to code anything I would have laughed at you so hard, it was like later in the year we're going to use the model that I wrote…so I like I shared that with them. [These things you are running into] are some of the frustrations that I experienced once… join the club. But I was like this is how proud I was what I figured it out, so I think that did help and they responded to that in answering [CT questions].

Though not necessarily directly related to the question, she decided to bring computational identity into the discussion after being prompted to reflect on CT. While not necessarily present in her feedback entries, Edna did work in the classroom to overcome identity issues that were averse to engaging in computation in her students. However, this prefacing could also indicate that she still is hesitant in pushing students to engage in CT *while also asking questions about CT Practice*. In fact, she pivoted to talk about one of her co-design partners, Parth, and his role in helping her design the questions:

> And there, there are times that [Parth] was like you should put it in a question about just specifically about the model, and I was like yeah that's a good idea, but I wouldn't have thought of it necessarily without that prodding, even though we had gone through those [demo] units where a lot of those types of questions were normal.

This shows that even though Edna was adept at addressing students’ computational practice in the classroom (as we saw in the previous chapter), she still saw herself as struggling to learn how to more formally assess CT practice. Edna then walked through her process of grading a summative assessment project from the same class, she followed basically the same set of practices: focusing in on elements of the project and looking

 © ISLS

for understanding of the underlying biological mechanisms at play. She often uses quizzes which provide her "a way of scanning through and saying okay, like…these people are getting these particular aspects of the science." When asked about her opinion on summative assessment for the course, Edna emphasized that she "preferred projects" and noted that "it was hard to make space for projects in the [CT] format." Notice that this implies that while the previously discussed SAQs are "summative" in style, Edna's idea of a summative assessment was much larger than a single question. In the assignment example Edna overviewed, students were given a number of options including creating a comic book about a topic in evolution. It is worth pointing out that this scale of creative outlet was not featured in her curriculum, indicating a possible disconnect in Edna's idea of assessment in her CT-ified classroom versus her traditional classroom.

Conclusions and next steps

In this work we see evidence that teachers engage can engage in deep reflection on how students engage with the computational activities in their units. However, through our analysis of the feedback teachers left for students, we see many opportunities to engage teachers in how they might adapt their existing feedback practices to new computational activities. By focusing in on a single teacher's feedback practice, we saw that while our teachers may be attempting to assess engagement in CT in their classrooms, there are still numerous difficulties to overcome to integrate their usual practices. While here we focused *existence*, in future work we plan to dive into the *content* of feedback to more specifically identify what dimensions of CT and STEM our teachers attend to in their practice. Linking this feedback to the in-classroom talk would give us a better understanding of the larger picture of frame alignment between CT and STEM in these classrooms. Finally, by analyzing assessment interviews across all our teacher subjects, we plan to tease out differences in feedback practices across these CT-ified units compared to these teachers' "traditional practice." These differences could then be used to construct additional tools to aid teachers in centering this practice in these new classrooms.

References

Bain, C., Dabholkar, S., & Wilensky, U. (2020). Confronting Frame Alignment in CT Infused STEM Classrooms. *CoolThink@ JC*, 91.

Bain, C., Hershokivtz, A., Dabholkar, S., Horn, M., & Wilensky, U. (2021, May). *Identifying Evidence of Student Engagement in CT via Automated Response Analysis*. National Association of Research in Science Teaching, Virtual.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *The Journal of Machine Learning Research*, *3*, 993–1022.

Brennan, K., & Resnick, M. (2012). New frameworks for studying and assessing the development of computational thinking. *Proceedings of the 2012 Annual Meeting of the American Educational Research Association, Vancouver, Canada*, *1*, 25.

Finzer, W. (2016). Common online data analysis platform (CODAP). *Emeryville, CA: The Concord Consortium.[Online: Concord. Org/Codap]*.

Gottipati, S., Shankararaman, V., & Lin, J. (2018). *Latent Dirichlet Allocation for textual student feedback analysis*.

Hoffman, M., Bach, F., & Blei, D. (2010). Online learning for latent dirichlet allocation. *Advances in Neural Information Processing Systems*, *23*, 856–864.

Kelter, J., Peel, A., Bain, C., Anton, G., Dabholkar, S., Horn, M. S., & Wilensky, U. (2021). Constructionist co-design: A dual approach to curriculum and professional development. *British Journal of Educational Technology*, *52*(3), 1043–1059. https://doi.org/10.1111/bjet.13084

Maloney, J., Resnick, M., Rusk, N., Silverman, B., & Eastmond, E. (2010). The scratch programming language and environment. *ACM Transactions on Computing Education (TOCE)*, *10*(4), 1–15.

Mouza, C., Yang, H., Pan, Y.-C., Ozden, S. Y., & Pollock, L. (2017). Resetting educational technology coursework for pre-service teachers: A computational thinking approach to the development of technological pedagogical content knowledge (TPACK). *Australasian Journal of Educational Technology*, *33*(3).

NGSS Lead States. (2013). *Next Generation Science Standards: For states, by states*. The National Academies Press.

Peel, A., Dabholkar, S., Anton, G., Wu, S., Wilensky, U., & Horn, M. (2020). *A case study of teacher professional growth through co-design and implementation of computationally enriched biology units*.

Rehurek, R., & Sojka, P. (2010). Software framework for topic modelling with large corpora. *In Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*.

Roschelle, J., Penuel, W., & Shechtman, N. (2006). *Co-design of innovations with teachers: Definition and dynamics.*

Sands, P., Yadav, A., & Good, J. (2018). Computational thinking in K-12: In-service teacher perceptions of computational thinking. In *Computational thinking in the STEM disciplines* (pp. 151–164). Springer.

Wallach, H. M. (2006). Topic modeling: Beyond bag-of-words. *Proceedings of the 23rd International Conference on Machine Learning*, 977–984.

Weintrop, D., Beheshti, E., Horn, M., Orton, K., Jona, K., Trouille, L., & Wilensky, U. (2015). Defining Computational Thinking for Mathematics and Science Classrooms. *Journal of Science Education and Technology*, *25*(1), 127–147. https://doi.org/10.1007/s10956-015-9581-5

Wilensky, U. (1999). *NetLogo* [Computer software]. Center for Connected Learning and Computer-Based Modeling. http://ccl.northwestern.edu/netlogo/

Windschitl, M., Thompson, J., Braaten, M., & Stroupe, D. (2012). Proposing a core set of instructional practices and tools for teachers of science. *Science Education*, *96*(5), 878–903.

Wu, S., Peel, A., Bain, C., Anton, G., Horn, M., & Wilensky, U. (2020). Workshops and Co-design Can Help Teachers Integrate Computational Thinking into Their K-12 STEM Classes. *Proceedings of International Conference on Computational Thinking Education 2020*.

Acknowledgements

This work was made possible through generous support from the National Science Foundation (CNS-1138461, CNS-1441041, DRL-1020101, DRL-1640201, and DRL-1842374) and the Spencer Foundation (#201600069).

International Society of the Learning Sciences

A Multimodal Study of Emotion Regulation in Student Dyads During Computational Model Building

Ashwin T S, Srigowri Mayasandra Prasanna, Joyce Fonteles, Gautam Biswas
ashwindixit9@gmail.com, srigowri.mayasandra.prasanna@vanderbilt.edu, joyce.h.fonteles@vanderbilt.edu, gautam.biswas@vanderbilt.edu
Vanderbilt University

Abstract: Collaborative problem solving (CPS) requires coordination of cognition and affect in complex, open-ended tasks. This study examines how pairs of high school students regulate learner-centered emotions (engagement, confusion, frustration, boredom, and delight) while constructing computational models in a STEM+C context. Extending Gross's (2015) process model to the group level, we analyze temporal dynamics of synchronized and unsynchronized emotional states during CPS. Multimodal data from six dyads (video, audio, system logs) were coded for CPS activities (e.g., information sharing, consensus building) and emotion transitions. Emotional alignment predominated, yet brief misalignments (especially between engagement and confusion) were frequent and often precipitated interpersonal regulation. Temporal analysis identified two dominant pathways: (1) transitions from misalignment to alignment that restore shared engagement, and (2) transitions between aligned states that shift partners from shared confusion to engagement. These pathways primarily occurred during information pooling and integration phases. Quantitative results show that effective regulation predicts higher model-building performance, with sustained engagement emerging as the stable endpoint of productive collaboration. The findings provide a temporal, group-level account of how students collectively manage emotion to sustain effective problem solving.

Introduction

Collaborative problem solving (CPS) in computational STEM integrates conceptual reasoning, coding, and social coordination. As students develop and test models, they share information, negotiate decisions, and experience emotions that shape collaboration quality (Cohn et al., 2025; Altaras et al., 2024). Learning-centered emotions (engagement, confusion, frustration, boredom, delight) fluctuate as groups construct solutions, face challenges, test hypotheses, and receive feedback. Managing these dynamics is crucial: positive emotions can strengthen persistence and cohesion, whereas negative emotions can disrupt communication and hinder joint reasoning. How groups handle emotion often determines whether collaboration remains productive (D'Mello et al., 2010; Haataja et al., 2022; Fonteles et al., 2024a; Ashwin et al., 2024; Akpanoko et al., 2024a; Fonteles et al., 2024b).

Emotion regulation research offers extensive insight into managing affect during learning. Gross's (2015) model distinguishes intrapersonal strategies for regulating one's own emotions from interpersonal strategies that influence others. Despite social contexts, much work in psychology and the learning sciences remains individual-focused. In socially shared and co-regulated learning (Järvelä et al., 2010; Hadwin et al., 2017), emotion is often treated as a background condition rather than a jointly managed process. Consequently, collective emotion regulation within CPS is not well understood. This gap is pronounced in computational modeling, where cognitive and emotional coordination are tightly coupled. Integrating science and computing concepts for model building depends on pooling information, negotiating understanding, and reaching consensus (Snyder et al., 2025). Emotions influence each activity: confusion can spark inquiry, frustration can motivate or impede progress, and delight can reinforce engagement. Yet prior work rarely examines group-level regulation—how partners detect emotional misalignments and co-regulate toward shared emotions that sustain collaboration (Chai et al., 2024; Altaras et al., 2024; Lönngren et al., 2024; Ashwin et al., 2025).

We address this gap by analyzing collaborative emotion regulation through emotional alignment and misalignment. We examine how pairs regulate emotions while solving computational modeling problems in a STEM+C environment. Using multimodal data—video, audio, and interaction logs—we track how partners' emotions align or diverge over time, how these shifts relate to CPS activities (information pooling, consensus building), and how regulation contributes to performance. We focus on two forms of collaborative regulation: (1) transitions from mismatched emotions to a shared state (misalignment-to-alignment) and (2) transitions from one shared emotion to another (aligned-to-aligned).

This study addresses three research questions:

- **RQ1:** Which emotional states are most frequently shared and misaligned during CPS?

 © ISLS

- **RQ2:** How do groups transition between aligned and misaligned states over time (frequency, typical paths, durations)?
- **RQ3:** How do aligned and misaligned intervals relate to CPS activities and subsequent performance?

Building on Gross (2015), we extend emotion regulation to group-level dynamics with a 2×2 framework distinguishing parallel and joint regulation within aligned and misaligned states. We map emotion trajectories onto CPS activities and performance, offering a temporally grounded account of collaborative emotion regulation in computational STEM learning. Together, the analyses show how emotions are collectively managed to sustain effective problem solving.

Background

Learning sciences research has established that emotion influences motivation, attention, and cognitive processing, impacting individual and group performance (Pekrun et al., 2017; Mänty et al., 2020; Ashwin and Guddeti, 2020; Desai et al., 2025; Yadannavar et al., 2024). In CPS, emotions arise both as internal reactions to cognitive demands and as socially situated responses to peers, group dynamics, and task progress. Positive emotions (e.g., engagement, delight) align with persistence, elaboration, and constructive dialogue, whereas frustration and boredom often co-occur with off-task behavior or disengagement (Bakhtiar et al., 2018; Ashwin et al., 2025). Certain negative emotions—particularly confusion—can be productive when regulated, prompting deeper reasoning or clarification (Isohätälä et al., 2020; Akpanoko et al., 2024b). Emotions thus fluctuate dynamically and continuously shape collaborative cognition.

Emotion regulation, notably Gross's process model (Gross, 2015), explains how individuals regulate and express emotions. Work in the learning sciences extends this framework to interpersonal and socially shared regulation, recognizing that learning often involves others. Research on co-regulation and socially shared regulation of learning (SSRL) (Järvelä et al., 2010; Hadwin et al., 2017) shows learners monitoring and managing their own and peers' cognitive and emotional states to maintain joint engagement, using reassurance, humor, and strategic reframing to sustain collaboration through affective challenges.

Despite progress, regulation is commonly treated as an individual phenomenon, with emotion peripheral to analysis and secondary to cognition. Studies typically rely on self-report or retrospective methods, capturing perceptions but not moment-to-moment evolution. Consequently, little is known about how regulation unfolds in real time within collaborative learning, especially when emotions are distributed and interdependent across partners. An underexplored area concerns learning-centered emotions—arising from cognitive demands—including engagement, confusion, frustration, boredom, and delight (D'Mello, 2012; D'Mello & Kory, 2015). These emotions shape epistemic processes, yet their collective regulation remains largely unexamined. While socio-emotional regulation studies identify strategies for managing climate and motivation, few investigate regulation of learning-centered emotions that fluctuates with conceptual understanding and task complexity during authentic collaboration. This gap is pronounced in computational STEM learning, where students coordinate conceptual science reasoning and computational thinking to build models. Such tasks demand ongoing negotiation of shared understanding, and resulting emotional dynamics directly influence collaboration.

Emotional misalignment (e.g., one partner frustrated, the other engaged) often requires regulation to restore balance; shared confusion, if effectively co-regulated, can transform into engagement and conceptual progress. Empirical analyses of how these configurations evolve during CPS—and how groups manage them—are scarce. Capturing these dynamics requires viewing regulation as both intrapersonal and as a temporal, relational process. Prior work frequently treats emotional states as discrete or aggregated measures, limiting insight into transitions—especially between aligned and misaligned configurations—and their links to regulation and performance. Moreover, most studies lack multimodal evidence synchronizing speech, action, and affect, obscuring the interplay among expression, regulation, and collaborative activity.

These limitations motivate a temporal, group-level perspective on emotion regulation. By examining how partners' emotions *align* or *diverge*—and how they regulate these dynamics in real time—research can move beyond individualistic models toward a comprehensive account of collective emotion regulation in learning. This study analyzes *collaborative emotion regulation* in authentic CPS tasks, using synchronized multimodal data to link emotional alignment and misalignment to activity patterns and performance. These dynamics point to two interrelated trajectories: resolution of misalignment and joint transformation of shared emotions. The next section formalizes these processes within a collaborative framework.

Conceptual framework

Social interactions during collaborative learning are complex and multi-dimensional, shaped by attention, discourse, emotional expression, and interpersonal coordination (e.g., Nasir et al., 2021; Vrzakova et al., 2020).

In collaborative problem solving (CPS), emotions are integral to how partners think, communicate, and sustain joint engagement. Emotional experiences can either support or disrupt collaboration (Mänty et al., 2020; Pekrun et al., 2017; Bakhtiar et al., 2018; Snyder et al., 2025). These emotional fluctuations are not random—they emerge in response to cognitive challenges, feedback, and social interaction within the group (Isohätälä et al., 2020). Thus, effective emotion regulation becomes essential to maintaining productive collaboration.

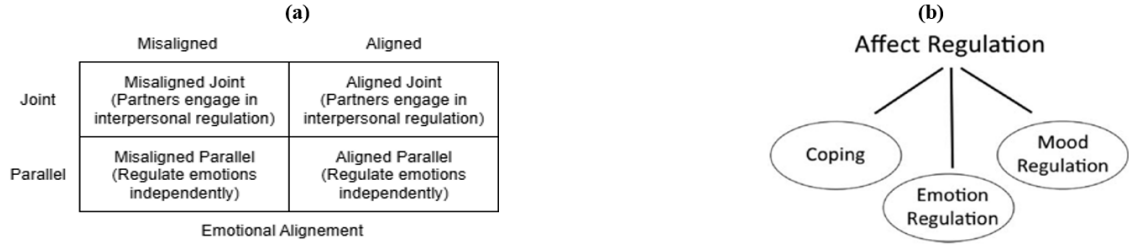
Gross's (2015) process model on emotion regulation conceptualizes regulation as the ways in which individuals influence which emotions they experience, when they experience them, and how these emotions are expressed. This model distinguishes between intrapersonal regulation, where an individual manages their own emotions, and interpersonal regulation, where one person intentionally influences another's emotions. Within collaborative learning, these processes are interwoven; learners must regulate both themselves and their partners while managing a shared cognitive and social space. Prior work on co-regulation and socially shared regulation of learning (SSRL) (Järvelä et al., 2010; Hadwin et al., 2017) has shown that such regulation is often negotiated collectively, particularly when groups encounter cognitive impasses or social tensions.

Building on this foundation, research on collaborative emotion regulation extends Gross's model, which describes the temporal stages of individual regulation (situation selection, attention deployment, cognitive change, and response modulation), to collaborative settings. In these more complex settings regulation must account for *emotional alignment* between partners. Alignment occurs when both group members experience similar emotional states (e.g., both engaged or both confused), whereas *misalignment* occurs when partners experience different emotions (e.g., one frustrated, one engaged). Collaborative regulation thus involves managing not only the intensity or valence of emotion, but also its distribution across individuals. To capture this relational nature of emotion regulation, we conceptualize dyadic emotional configurations within a 2×2 grid (Figure 1a). The x-axis represents emotional alignment, ranging from misaligned to aligned states, while the y-axis represents the regulatory focus, ranging from parallel to joint regulation.

This framework distinguishes four interactional configurations: (1) *Misaligned–Parallel*: Each student regulates emotions independently without explicitly engaging the partner's affect (e.g., one continues working while the other expresses frustration); (2) *Misaligned–Joint*: Partners recognize the emotional difference and engage in interpersonal regulation (e.g., one supports or reassures the other, restoring alignment); (3) *Aligned–Parallel*: Both students experience the same emotion but regulate independently (e.g., both quietly engaged or both confused without verbal coordination); and (4) *Aligned–Joint*: Both experience and regulate emotions together (e.g., jointly acknowledging confusion and co-regulating toward engagement). This grid highlights that emotional alignment alone does not guarantee effective regulation; rather, the quality of coordination—whether partners regulate in parallel or jointly—determines how successfully groups sustain or restore productive affective states. Transitions across these configurations capture the dynamic process of collaborative emotion regulation during CPS.

Figure 1

A 2×2 Framework and Hierarchical Model of Collaborative Emotion Regulation



(a): A 2×2 framework of collaborative emotion regulation during CPS. The horizontal axis represents emotional alignment (misaligned → aligned), and the vertical axis represents regulatory focus (parallel → joint), illustrating four possible dyadic configurations; **(b):** Hierarchical structure of affect regulation, adapted from Gross (2015).

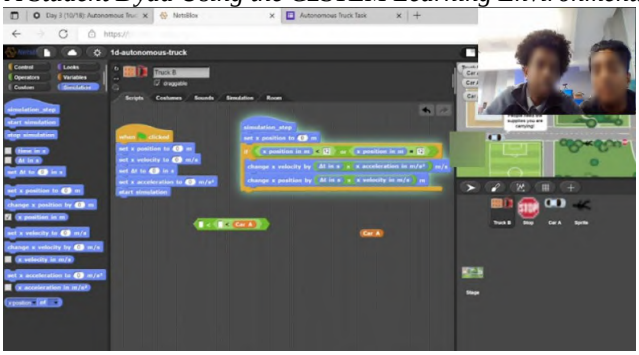
Figure 1b illustrates Gross's (2015) hierarchical framework of affect regulation, showing how coping, mood regulation, and emotion regulation relate conceptually. In this paper, we focus specifically on emotion regulation—and more precisely, on the regulation of *learning-centered emotions* (engagement, confusion, frustration, boredom, and delight) within collaborative problem solving. By mapping emotional alignment, misalignment, and transitions across time, we extend prior models of emotion regulation to collaborative contexts, offering a temporally grounded perspective on how emotions evolve and are collectively managed in group learning. In this paper, we analyze how transitions across these configurations—particularly from misaligned to aligned and from aligned to aligned—capture the dynamic processes of collaborative emotion regulation that emerge during CPS.

Learning environment and data

We conducted a two-month University Institutional Review Board (IRB) approved study with 10th-grade students at a public high school in the Southeastern United States. Weekly classroom sessions lasted two hours. Twenty-seven students participated, organized into 13 groups (12 dyads, one triad), each using a shared laptop for formative assessment and model-building in the Collaborative, Computational STEM (C2STEM) learning environment (Figure 2), where students built models to solve one-dimensional (1D) and two-dimensional (2D) challenges (Snyder et al., 2025). Students were free to collaborate without prior instruction and were encouraged to discuss and explore solution strategies together. All activity occurred on the shared laptop in class. We collected (a) pre- and post-test scores (summative assessments), (b) logs of computational model building from C2STEM, and (c) audio-video recordings captured with OBS on each group's laptop. Conversations were transcribed with Otter.ai and edited for accuracy by two researchers. Pre- and post-tests comprised four physics items (17 points) and three computing items (16 points). Scores were rubric-graded and reported as proportions of points earned over points possible. Groups lacking complete audio-video for both CPS tasks were removed, yielding a final analytic sample of six dyads with full recordings across both tasks.

Figure 2

A Student Dyad Using the C2STEM Learning Environment.



Method

Information pooling and consensus building

We analyzed group interactions via verbal exchanges on information pooling and consensus building. Each utterance was coded using Weinberger's (2006) framework for: (1) *Information Pooling (IP)* – eliciting and sharing relevant task information (a) *Elicitation:* one student asks the partner for task-related information (b) *Externalization:* a student articulates facts, observations, or describes actions within the environment; and (2) *Consensus Building* (CB) – jointly evaluating and integrating information (a) *Conflict-oriented*: disagreement or debate over interpretations or next steps; (b) *Integration-oriented*: building upon or integrating a partner's contribution; and (c) *Quick*: a proposal is accepted immediately without further discussion (Ashwin et al., 2025). An additional off-task label identified dialogue unrelated to the modeling activity. Two researchers independently coded all transcripts, achieving strong inter-rater reliability (Cohen's $\kappa = 0.77$). For each group, we computed the proportion of utterances per code relative to total utterances. These measures quantified information exchange and consensus processes and served as temporal references for segmenting episodes in later emotion analyses.

Affect analysis

We examined students' emotions during CPS using manual, time-aligned multimodal annotation. Data comprised synchronized screen recordings, webcam video (facial expressions, posture), audio, transcripts, and interaction logs. We focused on five learning-centered emotions—*confusion, frustration, delight, boredom*, and *engagement* (D'Mello, 2012; Pekrun et al., 2017; D'Mello & Graesser, 2012). Modalities provided complementary cues: facial and bodily signals indicated visible affect, while dialogue and tone validated context. When cues conflicted, the predominant signal guided labeling; explicit statements (e.g., "*this is frustrating*") overrode neutral expressions. Laughter or smiling after task success was coded as delight. Engagement was annotated as a neutral, task-focused state. Many students displayed steady gaze, neutral expressions, and consistent participation—behaviors aligned with epistemic emotions that support learning without strong positive or negative valence (D'Mello & Baker, 2010; Pekrun, 2012). This aligns with the notion of flow (Csikszentmihalyi, 1990), where learners exhibit deep immersion but limited external affect. Two trained annotators independently reviewed all sessions, recording for each episode the emotion label, start and end timestamps, and brief contextual justification. After iterative

 © ISLS

calibration, inter-rater reliability reached Cohen's $\kappa = 0.91$. Emotion timelines were synchronized across dyad members to identify aligned and misaligned intervals and subsequent regulation transitions.

Operationalizing emotional alignment and misalignment

Combining Gross's (2015) process model and socially shared regulation (Järvelä et al., 2010; Hadwin et al., 2017), we examined intrapersonal and interpersonal regulation. Intrapersonal regulation concerns managing one's own emotions; interpersonal regulation involves responding to or influencing a partner's affect. We defined emotional alignment as both dyad members experiencing the same emotion simultaneously (e.g., both confused or engaged) and misalignment as partners experiencing different emotions (e.g., one engaged, the other frustrated). Synchronized emotion timelines identified aligned and misaligned intervals.

We traced transitions between states (aligned → misaligned; misaligned → aligned) to characterize the temporal flow of collaborative regulation. Dialogue excerpts and contextual behaviors indicated how partners recognized and responded to affective differences, revealing efforts to negotiate and restore shared emotional states during CPS. Our focus was peer-to-peer regulatory dynamics emerging naturally in collaboration. We examined two transition types: (1) *misalignment-to-alignment*, where partners regulated toward a shared state, and (2) *aligned-to-aligned*, where partners jointly shifted from one shared emotion to another (e.g., confusion to engagement). Each transition was analyzed with CPS activity codes and performance to determine when and how regulation occurred. CPS segments defined task-based episodes (e.g., IP, CB phases), while emotional intervals captured continuous periods of alignment or misalignment within those segments; transitions marked the shifts between intervals.

Performance evaluation

Group performance was evaluated via rubric-based assessment of computational models in the C2STEM learning environment, covering physics and computing constructs: *Physics*: correct initialization of motion variables (position, velocity, acceleration) and appropriate relationships among them. *Computing*: simulation control, variable updating, and logical sequencing. Each criterion was scored 1 (present/correct) or 0 (absent/incorrect), totaling 19 points (11 physics, 8 computing). Scores were normalized across groups and tasks. Performance was analyzed alongside emotional-alignment data to test whether regulation, especially transitions restoring or maintaining alignment—was associated with stronger CPS outcomes.

Results and discussion

We analyzed 910 CPS segments from six student dyads over six hours of collaborative modeling. Each was coded for CPS phase and emotion. Overall, 107 segments (11.8%) showed emotional misalignment, with the rest aligned. There were 114 misaligned intervals, 65 of which involved Engagement–Confusion, and 107 regulation transitions. These findings formed the basis for further analysis of frequency, patterns, and performance links.

RQ1. Frequency and distribution of aligned and misaligned emotional states

During collaborative problem solving (CPS), students showed fluctuations among five emotions—engagement, confusion, frustration, boredom, and delight. About 12% of segments indicated emotional misalignment, while 88% showed alignment, mostly with engagement (≈57%) and confusion (≈24%), meaning students were often either focused or uncertain together. *Misalignment* was most common during Information Pooling (IP), especially Externalization, when students verbalized reasoning or explained code, revealing differing understanding or confidence. Consensus Building (CB), particularly Integration, also caused misalignment as partners debated ideas or negotiated solutions. Fewer misaligned segments occurred during Quick and Conflict phases, which were shorter and more decisive.

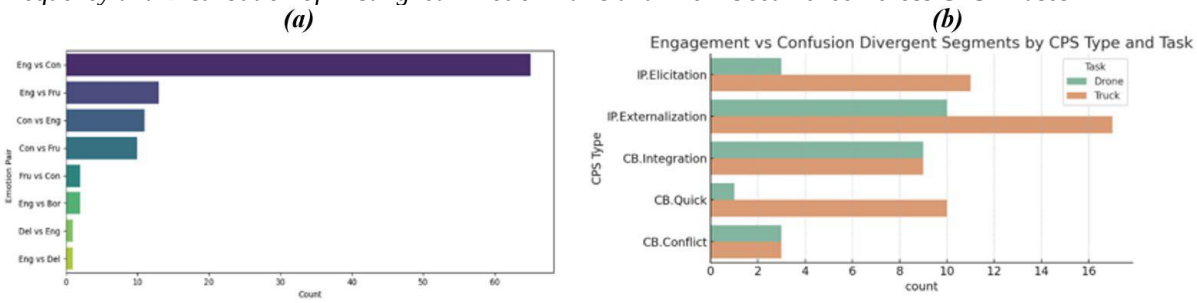
Dominant emotional pairings: Among misaligned intervals, Engagement ↔ Confusion was most common (≈ 65 instances; ~57%), followed by Engagement ↔ Frustration and Confusion ↔ Frustration (Figure 3a). These often reflected asymmetries—one partner focused, the other uncertain or challenged. Less common pairs (e.g., Engagement ↔ Boredom, Delight ↔ Engagement) occurred in under 2% of cases and are grouped as other low-frequency pairs. Since Engagement ↔ Confusion was most prevalent, we examined its distribution across CPS phases (Figure 3b). It appeared mostly during IP.Externalization and CB.Integration, indicating misalignments often coincided with elaboration or negotiation. When one partner scaffolded reasoning, misalignment frequently shifted to shared engagement, signaling successful regulation. Table 1 summarizes frequencies, durations, and CPS contexts. While aligned segments dominated total time, Engagement ↔ Confusion was a key misalignment often leading to regulation.

Statistical results: CPS phase significantly affected misalignment frequency ($F(4, 20) = 5.82$, $p = .003$, $\eta^2 = .54$), with IP.Externalization having more misalignments than CB.Conflict and CB.Quick. Segment durations varied across emotions, with Aligned–Confusion longer than Aligned–Engagement and Misaligned–Frustration.

Chi-square analysis indicated a highly uneven distribution of emotion pairs, mainly Engagement ↔ Confusion. Misalignment was a regular, meaningful part of CPS, indicating moments of one partner struggling while the other stayed focused, often leading to interpersonal regulation. Pooling highlighted differences; Integration helped repair alignment. Misalignment signals uneven understanding and triggers collaborative regulation to restore shared engagement.

Figure 3

Frequency and Distribution of Misaligned Emotion Pairs and Their Occurrence Across CPS Phases



(a): Frequency of misaligned emotion pairs. Engagement–Confusion dominated the dataset, followed by Engagement–Frustration and Confusion–Frustration; (b): Engagement-Confusion misalignment segments by CPS type and task

Table 1

Frequency and Duration of Aligned and Misaligned Emotional States across CPS Activities

Emotional Configuration	Frequency (%)	Duration (s)	Dominant Emotions	Typical CPS Phase	Example Regulatory Outcome
Aligned – Engagement	56.8	22 ± 5	Engagement	CB.Integration	Sustained joint focus
Aligned – Confusion	23.7	34 ± 7	Confusion	IP.Externalization	Shared uncertainty → clarification
Misaligned – Engagement ↔ Confusion	9.1	33 ± 9	Engagement / Confusion	IP.Elicitation → CB.Integration	Interpersonal repair → shared engagement
Misaligned – Engagement ↔ Frustration	2.7	28 ± 6	Engagement / Frustration	CB.Conflict	Brief tension, often resolved
Misaligned – Confusion ↔ Frustration	1.7	31 ± 8	Confusion / Frustration	IP.Externalization	Peer reassurance or withdrawal
Other Low-Frequency Pairs (a)	1.0	< 20	Eng ↔ Bored, Del ↔ Eng	Various	Transient misalignment

(a) "Other Low-Frequency Pairs" combine all misaligned configurations occurring < 2% each.

RQ2. Temporal transitions between aligned and misaligned states

Temporal analysis revealed patterns in how groups shifted between aligned and misaligned emotional states during CPS. As discussed, two main transition types were observed: misalignment-to-alignment and aligned-to-aligned. About 90% of misaligned episodes transitioned into an aligned state within the same task segment, often through explanations or reassurance (average 3.8 min, SD = 1.2). Shorter (~2 min) transitions occurred during CB.Conflict when tension was reduced. Longer sequences occurred during IP.Elicitation and early CB.Integration, where partners elaborated reasoning before re-establishing alignment. A smaller subset (~10%) involved partners moving from the same emotion to another shared state, most commonly Confusion → Engagement during CB.Integration, lasting about 5.0 min (SD = 1.5). Brief Engagement → Delight transitions

(~1.2 min) happened after successful tasks. Overall, 91% of sequences ended in shared engagement, confirming it as a stable outcome of CPS. Only eight episodes ended in sustained confusion, and none persisted in frustration or boredom beyond five minutes. Confusion and frustration episodes took about 10 min to recover, indicating greater regulatory difficulty.

Statistical results: A Kruskal–Wallis test showed a significant effect of CPS phase on regulation duration, $H(4) = 15.67$, $p < .001$, with longer transitions in IP.Elicitation and CB.Integration than in CB.Conflict. Transitions were not uniformly distributed across emotion pairings, $\chi^2(5, N = 107) = 11.64$, $p = .049$: Engagement ↔ Confusion was most likely to resolve to Engagement. The association between CPS phase and the resolution emotion was also significant, $\chi^2(4, N = 107) = 29.43$, $p < .001$, with Integration most likely to culminate in shared engagement. Table 2 summarizes durations (in minutes), CPS contexts, and characteristic mechanisms for the observed transitions. Most misalignments were brief and resolved through interpersonal regulation, while aligned transitions showed socially shared regulation deepening understanding. Engagement was the stable endpoint, reinforcing its role as the emotional basis of productive CPS. Longer transition durations in Integration suggest sustained emotional alignment is linked to conceptual co-construction, not passive convergence.

Table 2
Temporal Characteristics of Emotional Transitions During CPS

Transition	Starting State	Duration	Resolution	CPS Phase	Mechanism
Misaligned → Aligned	Engaged ↔ Confused	3.8 ± 1.2	Engagement	IP.Elicitation → CB.Conflict	Clarification / reassurance
Misaligned → Aligned	Engaged ↔ Frustrated	2.1 ± 0.7	Engagement	CB.Conflict	Supportive feedback
Aligned → Aligned	Confused → Engaged	5.0 ± 1.5	Engagement	CB.Integration	Joint cognitive reappraisal
Aligned → Aligned	Engaged → Delight	1.2 ± 0.5	Delight	CB.Quick	Shared success / closure

RQ3. Linking emotion regulation to CPS activities and collaborative performance

To relate regulation to collaborative quality, we linked aligned–misaligned transitions to CPS activities and Model performance was evaluated with a 19-point rubric, focusing on whether transitions restoring or maintaining alignment related to improved scores. Regulation occurred across all CPS phases, most often in IP and CB, with misalignments mainly in Elicitation and Externalization, where partners questioned and explained. Successful regulation in these phases restored alignment, fostering collaboration. Longer, stable aligned states appeared in CB.Integration, where groups refined models. Conflict and Quick episodes were brief, involving rapid affective stabilization. Regulated intervals correlated with higher performance, with a +3.0 point average gain.

Table 3
Regulation Contexts and Associated Performance

CPS Phase	Dominant Regulation	Duration (min)	Resolution Emotion	Δ score*
IP Elicitation	Misaligned → Aligned	3.5	Engagement	+1.2
CB Conflict	Misaligned → Aligned	2.0	Engagement	+0.8
CB Integration	Aligned → Aligned	5.0	Engagement → Delight	+3.0
CB Quick	Aligned → Aligned	1.0	Delight	+0.5

*Δ score = mean performance improvement relative to non-regulated intervals.

Regulation during Integration showed the strongest link, supported by a Spearman correlation of .63 ($p = .014$), especially in Integration ($r = .71$, $p < .01$) and Elicitation ($r = .52$, $p = .04$). No significant links were found for Conflict or Quick ($p > .10$). Episodes ending in shared engagement usually led to successful model iterations; those ending in shared confusion often stalled progress. A chi-square test showed that resolution

 © ISLS

emotions differed by performance tier ($\chi^2(3, N=107)= 12.76$, $p= .005$), with engagement-aligned intervals clustering in higher-rated teams. Table 3 details CPS phase, regulation types, durations, and their performance effects. Emotion regulation closely links to CPS activity. Extended regulation in Integration improved sustained engagement, reflection, and higher model scores. Brief or unresolved regulation in Elicitation or Conflict led to superficial alignment without deeper gains. Shared sustained engagement is key for productive collaboration and quality solutions.

Discussion

This study advances research on collaborative emotion regulation by situating it within learning-centered emotions during collaborative problem solving (CPS). We extend Gross's (2015) process model to collaborative regulation, showing alignment and misalignment as fluid configurations that evolve through cognitive and social coordination. Prior work on socially shared and co-regulated learning (e.g., Järvelä et al., 2010; Hadwin et al., 2017) links regulation to joint engagement but often treats affect as background to outcome. Our findings move beyond this view, offering fine-grained temporal evidence that emotions are dynamically regulated through interaction. Misalignments, especially engagement–confusion pairings were not failures but productive triggers for interpersonal regulation. Through talk, scaffolding, and reassurance, students restored shared engagement, positioning emotion regulation as a central mechanism for sustaining collaborative reasoning. These results bridge individual models of regulation and group-level coordination in CPS, showing how affective repair and progression underpin successful problem solving.

Beyond prior literature, our analyses yield several insights. First, emotion regulation in CPS operates through two complementary temporal pathways: emotional divergence inviting interpersonal repair (misalignment-to-alignment) and shared confusion evolving into collective engagement (aligned-to-aligned). Effective regulation thus depends not on constant positive affect, but on navigating emotional diversity. Second, the strong association between regulation during consensus-building phases and higher collaboration performance suggests sustained alignment supports deeper conceptual integration; alignment functions as a stabilizing force enabling cognitive progress. Third, our multimodal temporal approach—integrating emotion, discourse, and task phases—offers a methodological advance for capturing the microgenesis of regulation in situ, empirically validating emotion regulation as a collective, emergent phenomenon embedded in CPS. These results carry implications for emotion-aware learning in collaborative environments. They underscore the need to conceptualize emotion regulation as both social and epistemic, inseparable from joint meaning-making. For classroom and CSCL design, systems and instructors should attend to interpersonal affective configurations, i.e., moments of divergence or convergence, as key opportunities for scaffolding co-regulation. Automated detection of such transitions could inform adaptive prompts that foster emotional awareness and peer support. More broadly, the study supports a vision of collaboration in which emotional alignment is treated as a dynamic resource for maintaining shared focus, rather than a by-product of successful teamwork.

Limitations: Despite the paper's contributions, several limitations should be considered. The sample size, although adequate for microgenetic and qualitative temporal analysis, restricts our ability to generalize statistically. The setting—a single high school class working in a specific computational modeling environment—also limits the applicability of findings to other age groups or contexts. Furthermore, the emotion annotations, validated through multimodal evidence, depended on human interpretation, which introduces some subjectivity despite high inter-rater reliability.

Conclusions

This study provides empirical and theoretical insight into how students regulate their learning-centered emotions during collaborative problem solving. By analyzing temporal transitions between aligned and misaligned emotional states, we show that collaborative emotion regulation unfolds as an ongoing process of maintaining, repairing, and transforming shared affective engagement to advance shared understanding and problem-solving. Emotional misalignment especially *engagement–confusion pairings* emerged as a productive precursor to regulation, while sustained alignment during consensus-building phases predicted stronger collaboration performance. Together, these findings demonstrate that emotion regulation in CPS is not merely a reaction to affective disruption but a mechanism through which groups coordinate understanding and sustain joint inquiry. Extending Gross's (2015) model to the group level, this work positions emotional alignment as a fundamental dimension of collaboration dynamics, offering both conceptual grounding and methodological tools for studying emotion in learning. Future research should further investigate how emotion-aware designs can support students in recognizing and regulating these affective dynamics, enabling more resilient and productive forms of collaborative learning.

 © ISLS

References

Altaras, A., Nikitović, T., Mojović Zdravković, K., Krstić, K., Rajić, M., Pavlović Babić, D., & Jolić Marjanović, Z. (2024). The role of emotional intelligence in collaborative problem solving: A systematic review. Canadian Psychology/Psychologie canadienne.

Akpanoko, C. E., S, A. T., & Biswas, G. (2024). The Interplay of Affective States and Cognitive Processes in an Open-Ended Learning Environment: A Case Study. In Lindgren, R., Asino, T. I., Kyza, E. A., Looi, C. K., Keifert, D. T., & Suárez, E. (Eds.), Proceedings of the 18th International Conference of the Learning Sciences - ICLS 2024 (pp. 873-880). International Society of the Learning Sciences.

Akpanoko, C., Ashwin, T. S., Cordell, G., & Biswas, G. (2024, July). Investigating the relations between students' affective states and the coherence in their activities in open-ended learning environments. International Conference on Educational Data Mining.

Ashwin, T. S., Snyder, C., Prasanna, S. M., Mohammed, N., & Biswas, G. (2025, June). Using Markov Chain Analysis to Study the Relations Between Emotions, Cognitive Actions, and Performance in Collaborative Learning. In 18th International Conference on Computer-Supported Collaborative Learning (CSCL) 2025-CSCL Proceedings.

Ashwin, T. S., Snyder, C., Akpanoko, C. E., Srigowri, M. P., & Biswas, G. (2024, November). Combining multimodal analyses of students' emotional and cognitive states to understand their learning behaviors. In International Conference on Computers in Education.

Ashwin, T. S., & Guddeti, R. M. R. (2020). Impact of inquiry interventions on students in e-learning and classroom environments using affective computing framework: TS Ashwin, RMR Guddeti. *User Modeling and User-Adapted Interaction*, *30*(5), 759-801.

Baker, R.S., D'Mello, S.K., Rodrigo, M.M.T., Graesser, A.C.: Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive-affective states during interactions with three different computer-based learning environments. International Journal of Human-Computer Studies 68(4), 223–241 (2010)

Bakhtiar, A., Webster, E. A., & Hadwin, A. F. (2018). Regulation and socio-emotional interactions in a positive and a negative group climate. Metacognition and learning, 13(1), 57-90.

Chai, H., Hu, T., & Wu, L. (2024). Computer-based assessment of collaborative problem solving skills: A systematic review of empirical research. Educational Research Review, 43, 100591.

Cohn, C., Snyder, C., Fonteles, J. H., TS, A., Montenegro, J., & Biswas, G. (2025). A multimodal approach to support teacher, researcher and AI collaboration in STEM+ C learning environments. British Journal of Educational Technology, 56(2), 595-620.

Csikszentmihalyi, M. (1990). Literacy and intrinsic motivation. Daedalus, 115-140.

Desai, A., N_V_J_K, K., Gupta, P., Vanahalli, M., Rajendran, R., TS, A., & Biswas, G. (2025). Exploring Emotional Trajectories through Valence and Arousal in Open-Ended Math Learning Environments.

D'Mello, S. K., Lehman, B., & Person, N. (2010). Monitoring affect states during effortful problem solving activities. International Journal of Artificial Intelligence in Education, 20(4), 361-389.

D'Mello, S. (2012). Monitoring affective trajectories during complex learning. In Encyclopedia of the Sciences of Learning (pp. 2325-2328). Springer, Boston, MA.

D'Mello, S., & Graesser, A. (2012). Dynamics of affective states during complex learning. Learning and Instruction, 22(2), 145-157.

D'Mello, S. K., & Kory, J. (2015). A review and meta-analysis of multimodal affect detection systems. ACM computing surveys (CSUR), 47(3), 1-36.

Fonteles, J., Davalos, E., Ashwin, T. S., Zhang, Y., Zhou, M., Ayalon, E., ... & Biswas, G. (2024, July). A first step in using machine learning methods to enhance interaction analysis for embodied learning environments. In International Conference on Artificial Intelligence in Education (pp. 3-16). Cham: Springer Nature Switzerland.

Fonteles, J., Srivastava, N., Davalos, E., TS, A., & Biswas, G. (2024, November). Designing an AI-Enhanced Timeline for Monitoring Multimodal Interactions in Embodied Learning Environments. In International Conference on Computers in Education.

Gross, J. J. (2015). Emotion regulation: Current status and future prospects. Psychological inquiry, 26(1), 1-26.

Hadwin, A., Järvelä, S., & Miller, M. (2017). Self-regulation, co-regulation, and shared regulation in collaborative learning environments. In Handbook of self-regulation of learning and performance (pp. 83-106). Routledge.

Haataja, E., Malmberg, J., Dindar, M., & Järvelä, S. (2022). The pivotal role of monitoring for collaborative problem solving seen in interaction, performance, and interpersonal physiology. Metacognition and Learning, 17(1), 241-268.

Isohätälä, Jaana, Piia Näykki, and Sanna Järvelä. "Cognitive and socio-emotional interaction in collaborative learning: Exploring fluctuations in students' participation." Scandinavian Journal of Educational Research 64, no. 6 (2020): 831-851.

Järvelä, S., Volet, S., & Järvenoja, H. (2010). Research on motivation in collaborative learning: Moving beyond the cognitive–situative divide and combining individual and social processes. Educational psychologist, 45(1), 15-27.

Lönngren, J., Bellocchi, A., Berge, M., Bøgelund, P., Direito, I., Huff, J. L., ... & Tormey, R. (2024). Emotions in engineering education: A configurative meta-synthesis systematic review. Journal of Engineering Education, 113(4), 1287-1326.

Mänty, K., Järvenoja, H., & Törmänen, T. (2020). Socio-emotional interaction in collaborative learning: Combining individual emotional experiences and group-level emotion regulation. International Journal of Educational Research, 102, 101589.

Nasir, J., Kothiyal, A., Bruno, B., & Dillenbourg, P. (2021). Many are the ways to learn identifying multi-modal behavioral profiles of collaborative learning in constructivist activities. International Journal of Computer-Supported Collaborative Learning, 16(4), 485-523.

Pekrun, R., Linnenbrink-Garcia, L.: Academic emotions and student engagement. Handbook of research on student engagement, 259–282 (2012)

Pekrun, R., Lichtenfeld, S., Marsh, H. W., Murayama, K., & Goetz, T. (2017). Achievement emotions and academic performance: Longitudinal models of reciprocal effects. Child development, 88(5), 1653-1670.

Pekrun, R. (2024). Control-value theory: From achievement emotion to a general theory of human emotions. Educational Psychology Review, 36(3), 83.

Snyder, C., Cohn, C., Fonteles, J. H., & Biswas, G. (2025). Using collaborative interactivity metrics to analyze students' problem-solving behaviors during STEM+ C computational modeling tasks. Learning and Individual Differences, 121, 102724.

Snyder, C., Wen, C. T., Hutchins, N. M., Vatral, C., Liu, C. C., & Biswas, G. (2024). Investigating collaborative problem solving behaviors during stem+ c learning in groups with different prior knowledge distributions. In Proceedings of the 17th International Conference on Computer-Supported Collaborative Learning-CSCL 2024, pp. 107-114. International Society of the Learning Sciences.

Vrzakova, H., Amon, M. J., Stewart, A., Duran, N. D., & D'Mello, S. K. (2020, March). Focused or stuck together: Multimodal patterns reveal triads' performance in collaborative problem solving. In Proceedings of the tenth international conference on learning analytics & knowledge (pp. 295-304).

Yadannavar, S., Ashwin, T. S., Pathan, R., Vanahalli, M. K., & Rajendran, R. (2024, December). Multimodal analysis of learning-centered emotions and cognitive processes in open-ended learning environments. In 2024 IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE) (pp. 1-8). IEEE.

Acknowledgments

This research was supported by the National Science Foundation AI Institute Grant No. DRL-2112635. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

International Society of the Learning Sciences

Exploring Cognitive, Social, and Metacognitive Dimensions of Student-Student-LLM Collaboration in Problem Solving

Victoria Delaney, San Diego State University, vldelaney@sdsu.edu
Ibrahim Oluwajoba Adisa, Stanford University, jobaa@stanford.edu

Abstract: If large language models (LLMs) are to be framed as collaborative tools, it is important to understand the nature of LLM collaboration and how it may differ from human collaboration. To this end, we studied problem solving episodes between undergraduate dyads and ChatGPT, observing the extent to which students' interactions with ChatGPT mediated their cognition, social interactions, and/or metacognition. We found mixed results among dyads (including one that avoided ChatGPT altogether) and elaborate on four cases in-depth, highlighting interactions between students and juxtaposing them with ChatGPT collaboration. Dyads that collaborated with ChatGPT spent more talk turns monitoring its output whereas dyads who collaborated with each other developed more cognitive and metacognitive awareness of the task. Our findings raise implications for tradeoffs that educators should consider when deciding if and when to introduce LLMs into problem solving.

Introduction

As artificial intelligence (AI) systems evolve, their growing ability to reason, hold social conversations, and offer feedback has sparked questions about whether AI can serve as a true collaborator (Scardamalia & Bereiter, 2020). Large language models (LLMs) in particular demonstrate interactive capabilities that allow them to engage in forms of reasoning and discourse once thought to be uniquely human. These advances invite us to reconsider what collaboration means when one partner in a group configuration is a LLM: what kinds of reasoning or social exchanges can such collaboration enable, and where might its limits lie? Effective collaboration requires participants to establish joint attention, monitor reasoning, and build on shared epistemic knowledge (Barron, 2003). Dialogue and explanation play central roles in supporting both cognitive and metacognitive engagement, allowing partners to co-construct knowledge rather than simply divide and conquer.

We examine collaboration in this study through mathematical problem solving, a form of activity with a long and rich history of technology-mediated collaboration (Kaput, 1992). In mathematics, collaboration has long been viewed as a productive context for developing conceptual understanding and problem-solving competence. When students explain their reasoning, question one another's assumptions, or compare alternative strategies, they not only expand their mathematics schemas to include new ideas, but also revise and reconfigure existing ideas (Boaler, 2016; Goos et al., 2002; Slavin, 2014). These social exchanges support the kind of metacognitive monitoring that helps learners evaluate not just what they know, but how they know it. However, it is unclear the extent to which collaborating with an AI tool provides similar cognitive and metacognitive restructuring, as the social dimensions of collaborating with a human versus a pre-trained machine differ. Students must decide when and how to invite the LLM to contribute to their reasoning, interpret its contributions, and evaluate its suggestions against their own ideas. These interactions may also shift how students share cognitive responsibilities—whether they rely more on the AI's responses, or on one another's situated interpretations and epistemic judgments of mathematics problems. We argue that understanding these potential modes of difference is especially relevant as students begin to take up LLMs in classrooms, and raise important questions for enacting instruction with AI. For instance, if multiple students and one LLM interact, to whom is authority conferred? How might differences in cognitive, social, and affective capacities between humans and machines shape group processes and outcomes? (Rizos et al., 2025; Bastani et al., 2025).

To address questions like this, we investigated how collaboration unfolds when undergraduate dyads solved a nonroutine function analysis task and ChatGPT was available. We examine the ways in which cognitive, social, and metacognitive processes are distributed across students and ChatGPT and how these processes influenced the degree to which students viewed ChatGPT as a collaborator. Through this analysis, we aim to contribute to ongoing discussions about how LLMs can be integrated into problem solving in ways that enrich (not replace) social and cognitive dimensions of human interaction. We ask: *How do students collaborate with (a) each other and (b) LLMs during mathematical problem solving?*

Collaborative problem solving

Collaboration is generally defined as a social activity that involves individuals engaging in active discourse and contributing meaningfully within a group (Roschelle & Teasley, 1997; Ouyang et al., 2023). During collaboration,

 © ISLS

team members share information, make decisions, and implement solutions they mutually agree upon (Sun et al., 2020). Collaborative problem solving (CPS) combines the complexities of collaboration and problem solving into a dynamic and mutually beneficial process (Swiecki et al., 2020). Because collaboration is as social as it is cognitive, successful collaboration depends on effective communication among partners and requires trust, respect, and openness to each member's contributions and intentions. CPS allows students to bring together diverse perspectives, experiences, and areas of expertise to address a shared problem or goal.

Several CPS frameworks suggest that collaboration involves a complex interplay of social, cognitive, and metacognitive skills (An et al., 2024; Andrews-Todd & Forsyth, 2020; Hesse et al., 2015). Social skills such as participation, maintaining shared understanding, perspective-taking, and attentive listening are critical for effective group coordination. Through these interactions, group members keep one another informed about progress, express ideas or concerns, and adjust their tasks and strategies accordingly (An et al., 2024). Cognitive skills, including planning, ideating, and monitoring, enable students to reason through problems, integrate multiple perspectives, and construct shared solutions (OECD, 2017). Metacognitive skills play an equally important role, as they reflect students' ability to monitor and regulate both their cognitive activities and social interactions during collaboration (Cini et al., 2023; Dindar et al., 2020). The interplay among these three dimensions not only influences individual performance but also shapes the overall quality and productivity of group outcomes (Goos et al., 2002). These interconnections are so critical that recent studies have shown how the diversity of individuals' social and cognitive skills can determine the success of collaboration. For example, Li et al. (2022) found that collaborative outcomes improved when students maintained active social interactions and made substantive cognitive contributions. Similarly, Andrews-Todd and Forsyth (2020) reported that in triadic settings, the presence of at least one member with strong social and cognitive abilities improved CPS.

As collaborative interactions become increasingly mediated by LLMs, it is likely that the cognitive, social, and metacognitive dynamics within teams will evolve in new ways. LLMs, trained on vast datasets, are likely to contribute more strongly to cognitive processes such as information retrieval, reasoning, and problem decomposition. In contrast, human collaborators bring contextual and situated awareness, lived experiences, and sophisticated social-emotional skills that support coordination and joint meaning-making (Scardamalia & Bereiter, 2020). Understanding how these distinct capabilities interact is essential for supporting human–LLM collaboration, especially in multi-human settings where the balance of cognitive and social regulation may shift.

Human-LLM collaboration

Human–LLM collaborative problem solving refers to the joint effort between humans and AI systems to solve problems and produce creative outputs (Davis, 2013). As LLMs like ChatGPT gain popularity for their humanistic capabilities, human–LLM collaboration is becoming prevalent across domains, including education. Like other forms of collaboration, successful outcomes depend on good coordination between partners (An et al., 2024): sharing information, responding substantively to each other's contributions, and jointly advancing the task (Scardamalia & Bereiter, 2020). Rezwana and Ford (2025) characterize this purposeful and bidirectional exchange of information between humans and AI as AI communication, emphasizing that effective collaboration occurs when AI systems can interact and offer feedback in ways that resemble human teammates (Mamykina et al., 2002). In the context of LLMs, prompts serve as the primary means of communication and act as a window into individuals' cognitive and metacognitive engagement with their AI partners (Subramonyam et al., 2024). Through prompting, humans articulate their goals, revise their input, and co-construct ideas with the AI system. The specificity of these prompts and revisions often determines the quality of the collaboration.

Research on human–LLM collaboration in education is still emerging, and recent studies report mixed findings. For example, bidirectional exchanges of ideas between humans and AI have been reported to create opportunities for novelty (Urmeneta & Romero, 2024), scaffold learners' knowledge (Bastani et al., 2025), and enhance creativity in some instances (Huang et al., 2023). However, it remains unclear how best to structure these collaborations, particularly when students are still learning how to solve mathematics problems and are developing skill, confidence, and fluency in mathematics. For instance, Bastani et al. (2025) implemented a randomized control trial and found that students who used an AI system while solving problems achieved higher test scores during collaboration but performed worse once AI was no longer available, compared to peers who had never used AI. These findings suggest that while AI or LLM collaboration can improve problem solving accuracy, it may not necessarily strengthen humans' abilities to develop independent skill in problem solving.

In addition to communication, several factors, such as team members' prior experiences, perceptions, and levels of trust in AI systems, can influence how they collaborate with AI and the overall quality of that collaboration. Research indicates that individuals' beliefs about AI competence, reliability, and intent strongly shape their engagement behaviors and reliance patterns during collaborative tasks (Muhammad et al., 2024; Adisa & Adefisayo, 2025). These factors highlight the importance of studying not only what AI contributes to problem

solving but also how humans interpret, negotiate, and regulate their interactions with AI partners. Our study extends these conversations by examining how such collaborative dynamics unfold when students work with a LLM in dyads to solve open-ended mathematical problems.

Methods

The purpose of our broader exploratory work was to investigate, in detail, how students think and reason through mathematics tasks when a LLM is made available to them. We conducted this work during spring-fall 2025. The overarching study included 10 individuals solving with only ChatGPT-4o, 10 individuals solving with ChatGPT and the internet, and 16 students who collaborated with each other in addition to having ChatGPT available. This paper reports on the third condition. We recruited students via physical campus advertisements. Students needed to have taken a mathematics course in the past year to reduce the likelihood that they would be unfamiliar with the tasks' mathematics concepts, which we asked them about on a screening survey. We also asked students if they would be willing to bring a friend to the session, and eight participants agreed (16 collaborative participants total). Students usually brought a friend in the same major, but two did not (Table 1).

Table 1

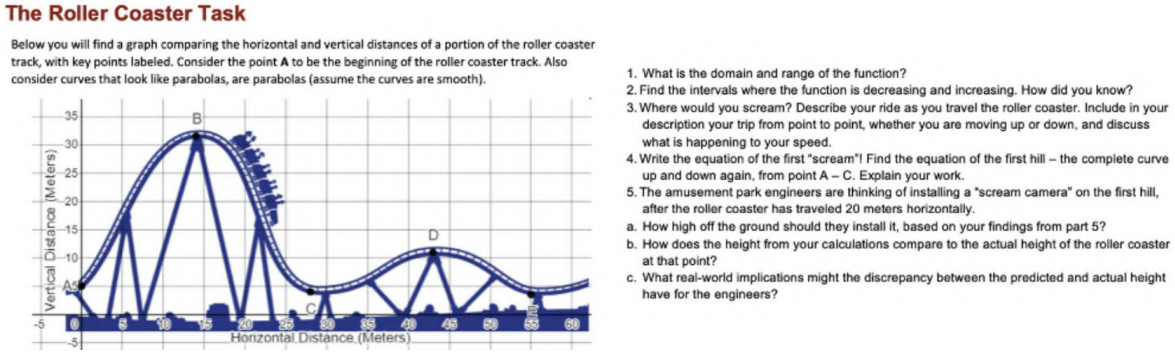
Summary of Study Dyads

Student Pseudonyms	Year	Major	Number of Math Courses
*Alex and Kam	Both senior	Both math teaching	Both: 15
*Elsa and Gretchen	Both junior	Both business administration	Both: 4
Celia and Emma	Both junior	Emma: Child development Celia: Liberal studies	C: 4; E: 1
Itzy and Milan	Both senior	Both mathematics	I: 17; M: 16
Ken and Jazlyn	Both freshmen	Both civil engineering	K: 5; J: 4
*Connor and Colin	Both freshmen	Both mechanical engineering	Con: 4; Col: 5
*Jaiden and Natalia	Both senior	Both math teaching	Both: 14
Rich and Junho	Both senior	Rich: Liberal studies Junho: Cognitive science	R: 10; J: 9

*Indicates Case Study

Figure 1

The Roller Coaster Task, Adapted From the Virginia Department of Education



Recruited students participated in a 90-minute study involving think-aloud, task-based, cognitive interviews (diSessa, 2007). Each student was first interviewed individually by a researcher, who asked questions about students' typical problem solving strategies, AI uses (in math, school, and in daily life), AI fairness, and the degree to which their instructors encouraged AI use. Researchers followed up with additional probing questions until they fully understood each participant's perspective. Then the participants were brought back into the same space and one researcher assumed the role of lead interviewer. They facilitated three 20-minute tasks: algebra, statistics, and calculus, which we acquired from publicly available sources and modified for achievability within the time frame. Due to page constraints, we focus on the first task: Roller Coaster (Figure 1), because students appeared to have the most consistent prior (algebra) knowledge on this problem. The written task (Remillard, 2005) asked students to analyze a "roller coaster-like" mathematics function and derive a mathematical model from it. We allowed students to interpret its meaning however they wanted; for instance, Connor and Colin interpreted it as a physics problem dealing with potential and kinetic energy. We did not consider their answers to be "wrong" if they did not use parabolas, as written in the task (some groups justified the accuracy of a sine

 © ISLS

wave). At no point did we require students to use ChatGPT; it was there if they determined a need to interact with it. A second researcher managed video, audio, and iPad screen recordings of student work, and took detailed field notes on (1) students' problem solving strategies, (2) student collaboration with ChatGPT, and (3) the purpose or rationale for which they appeared to collaborate with it. Once all three tasks were completed, each student was interviewed individually to debrief their experience.

Data analysis

Data analysis was iterative and partially informed by an adjacent analysis (Delaney et al., under review), where we examined collaboration patterns and AI trust among four dyads in this data corpus. However, that study did not account for pro-social collaboration moves that ChatGPT could have taken and thus, tended to frame AI as a tool rather than a humanistic collaborator. In this analysis, we do not make such *a priori* assumptions. We use an analytical framework from Goos et al. (2002), based on Vygotskian sociocultural theory, that characterizes mathematics collaboration in three ways: *self-disclosure* (statements about one's own understanding), *other-monitoring* (statements about one's understanding of a groupmate's understanding) or *feedback request* (a request for feedback on one's own thinking). We hypothesized that the degree to which students viewed ChatGPT as their collaborator could be observed in Goos' collaboration moves; for instance, pro-LLM collaboration could include students asking ChatGPT for feedback on their ideas, critiquing its output against their own ideas, or using its output as the basis of new ideas or problem solving strategies. We also propose that these pro-social mathematical collaboration moves reflect broader notions of collaboration, such as attentive listening, shared understanding, and perspective taking introduced earlier in our theoretical framework.

Both authors individually reviewed the data corpus: eight autotranscribed then cleaned (by hand) transcripts of students' collaboration on the roller coaster task, ChatGPT logs for each session, and video recordings during moments where students' crosstalk, gesturing, and interactions were difficult to interpret from written transcripts alone. We met and discussed patterns, realizing that some students tended to treat ChatGPT more like a tool and others more like a collaborator. We then analyzed the data by applying Goos et al.'s framework at the level of conversational turn-of-talk, attending carefully to relationships between students and ChatGPT (Jordan & Henderson, 1995). We applied one code per talk turn. If a talk turn appeared to contain multiple codes (e.g., one self-disclosure followed by a request for feedback), we split it into two talk turns and applied codes separately. Not all talk turns were coded (e.g., "Okay, let's read the first task."). After reaching an initial Cohen's kappa coefficient of 0.86 on the first two transcripts, the first author coded the rest. The research team met two more times to discuss codes. As patterns emerged, we realized that students incorporated ChatGPT in three ways: as a collaborator, as a peripheral teacher or tutor, or as a tool. We wrote descriptive cases for each collaboration pattern (Yin, 2017). These cases form the basis of our findings.

Findings

Of the eight dyads, three groups (Alex/Kam, Connor/Colin, and Itzy/Milan) treated ChatGPT like an information-seeking tool, three groups (Ken/Jazlyn, Rich/Junho, and Elsa/Gretchen) treated ChatGPT like a peripheral teacher or tutor, and two groups (Jaiden/Natalia, Emma/Celia) treated ChatGPT like a collaborator. Groups who used ChatGPT like a tool used it only to extract information (e.g., "what is the equation of a parabola in vertex form?"). Groups who used ChatGPT like a peripheral teacher or tutor used it to extract information and for explanations about how to solve processes (e.g., "how do you find the domain and range of a function?"). Groups who used ChatGPT as a collaborator asked it questions beyond information and explanation, instead asking it for input in prosocial ways as outlined by Goos et al. (2002). We illustrate a case of each and one dyad who decided not to use ChatGPT in their problem solving process.

Alex and Kam: ChatGPT is a tool

Alex and Kam collaborated almost entirely with one another. Among the five portions of the task, they tended to individually self-disclose what they each knew and understood about the problem first, then build off of one another's ideas (other monitoring), then self-disclose how they made sense of their combined solution strategy. Figure 2 (left) shows these collaboration patterns. Rather than solicit each other's feedback (Figure 2, right), they tended to pose ideas (self-disclosure) and then critique them using their knowledge of mathematics (other monitoring), as shown in this interaction in task part 3, describing their ride:

> Alex (10:16): You're going up a little bit. [self-disclosure]
>
> Kam (10:17) And then it starts to flatten out. [self-disclosure]
>
> Alex (10:20): Yeah, our slope isn't our speed, because it's not in time. [other monitoring]
>
> Kam (10:25): No, but it's not linear. [other monitoring].

 © ISLS

Alex (10:27): Yeah, but our x-axis isn't time, so it's not distance over time. [other monitoring]

After several rounds of negotiating what the roller coaster would do at various points on the track (via other monitoring), Alex and Kam reached agreement that the roller coaster's speed was relative to its position, although each offered their own explanations:

Alex (10:41): …and so, because it goes up more in the beginning, you're going faster. And then as you get closer to B, it slows down. [self-disclosure]

Kam (11:03): You're definitely moving faster than you are between C and D, or D and E. [self-disclosure]

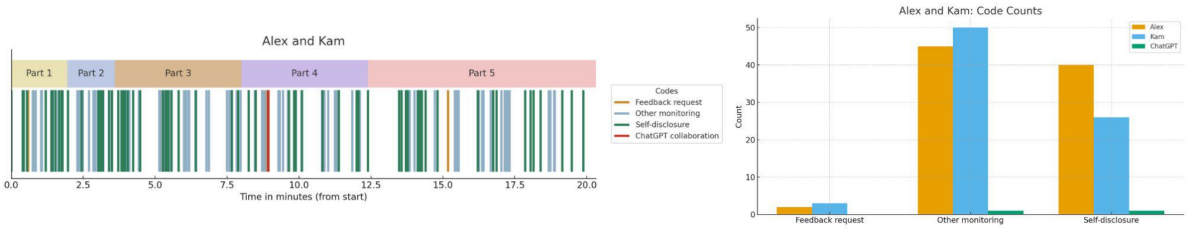
Alex and Kam only prompted ChatGPT once, during task part 4, as they were attempting to remember how to create an equation in vertex form for a parabolic function:

Alex (12:47): So our vertex, I haven't created an equation for a parabola in a minute. It's going to be a negative slope for sure. [self-disclosure]

Alex (13:00): Wait, we should ask the chatbot how do you create an equation for a parabola. [other monitoring].

Kam (13:05): Okay. (*prompts ChatGPT: "how do you create your equation for a parabola using your vertex and two other points"*)

Figure 2
Alex and Kam's Problem Solving with ChatGPT



Collaboration codes and ChatGPT interactions over time (left) and collaboration frequencies (right).

In response to their prompt, ChatGPT self-disclosed information about the general form of a parabola in vertex form, gave directions on how to plug in points to determine the "a" coefficient (the parameter that determines the width of the parabola), and gave a worked example. Alex and Kam, however, only viewed ChatGPT's overview of how to solve for the equation before generating an idea of their own. Upon seeing the reminder to "solve for a" in ChatGPT's directions, Alex reflected, "I don't know how to find the "a," but actually, I feel like we could find the "a," because we just need to plug in one of the other points with that (the vertex) as a variable." Alex and Kam then pursued their own solution strategy and found "a" using the vertex (14, 32) and the point (0, 5). Once immersed in their strategy, they did not appear to remember that ChatGPT gave directions on how to find "a" until the very end, when Alex remarked, "Oh, we didn't even look at [ChatGPT's worked example]...and looking at what ChatGPT gave us, it did give us exactly what we did. But we did come up with it on our own without even looking at what [it] gave us." ChatGPT thus functioned as a tool that provided Alex and Kam with the means to confirm their method without doing much cognitive lifting.

Elsa and gretchen: ChatGPT is a peripheral tutor

Elsa and Gretchen collaborated with one another until they encountered a conceptual or procedural roadblock, then used ChatGPT to become unstuck (see Figure 3). They tended to give their own mathematical ideas through self-disclosure but, in contrast to Alex and Kam, did not monitor each other's thinking in ways that surfaced contradictions. Rather, they tended to build from each other without discussion. During instances of other monitoring, they tended to respond to each other's sensemaking about the problem and whether they had enough information to solve it, as opposed to conversing with each other to build confidence that they both understood the meaning of the solution they were collectively constructing. We illustrate the portion of the task where they constructed intervals of increasing and decreasing behavior:

Elsa (1:42): (*reading from task*) What are the intervals where the function is increasing or decreasing, how do you know?

Gretchen (1:52): Okay so, increasing is zero to 32. And then it looks like, 27. [self-disclosure]

Elsa (2:10): Yeah 27, maybe 28. That's pretty close to 30. [other monitoring]

Gretchen (2:12): 28 to 40? [self-disclosure]

Elsa (2:15): 44? [self-disclosure]

Gretchen (2:19): *writes down 44*.

Here, Elsa and Gretchen do not attempt to understand or discuss each other's estimates for the intervals of increasing and decreasing, but rather build off of each other's estimates by posing their own ideas. Much of their collaboration took this form, except during instances where neither partner had an idea about what to do next. In these cases, Elsa and Gretchen drew upon ChatGPT. They explained that they used ChatGPT in instances where the problem did not provide enough information (as opposed to, they would be responsible for holding the information in their prior knowledge). Elsa explained that they treated ChatGPT like a teacher that they would call upon if they were stuck or stumped in class: "There's not enough information [in the task], if we had a professor, we'd ask the professor. But we're asking [*ChatGPT*], this is our professor right now." We illustrate their ChatGPT collaboration in task part 4, as they solved for the equation of the first scream:

Elsa (5:49): Write the equation of the first scream. What is the equation? Is there a formula? [self-disclosure]

Gretchen (5:56): I feel like there is a formula for…parabolas? [self-disclosure]

Elsa (6:08): *prompts ChatGPT "formula for parabolas"* [other monitoring].

Elsa (6:30): …it says y is equal to $ax^2 + bx + c$. Is this the formula? [other monitoring]

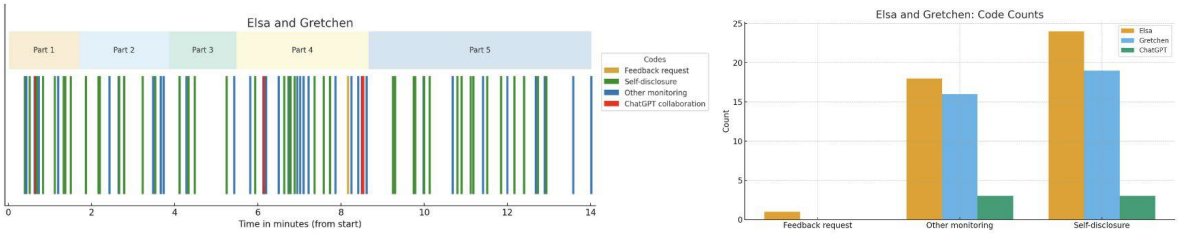
Gretchen (6:38): think our C is five. [self-disclosure]

Elsa (6:44): I think I've seen this before. [self-disclosure]

Gretchen (6:46): Because that's where it starts. [self-disclosure]

Figure 3

Elsa and Gretchen's Problem Solving with ChatGPT



Collaboration codes and ChatGPT interactions over time (left) and collaboration frequencies (right).

ChatGPT explained the components of a parabolic equation in standard and vertex form. Elsa and Gretchen appeared to notice standard form first and, perhaps because Elsa had seen it before, used it as the basis of their solution. They took up ChatGPT's request for help: "Want help graphing one or finding the vertex/focus/axis of symmetry?" and identified three points from the task to input into ChatGPT. They then determined that those three points formed the basis of an equation ("the formula") that represented their solution - the equation of the hill where the first scream occurred. They adopted ChatGPT's output: $y = 5x^2 + 32x + 5$, as their solution. In this case, ChatGPT's collaboration consisted of solving the problem for Elsa and Gretchen.

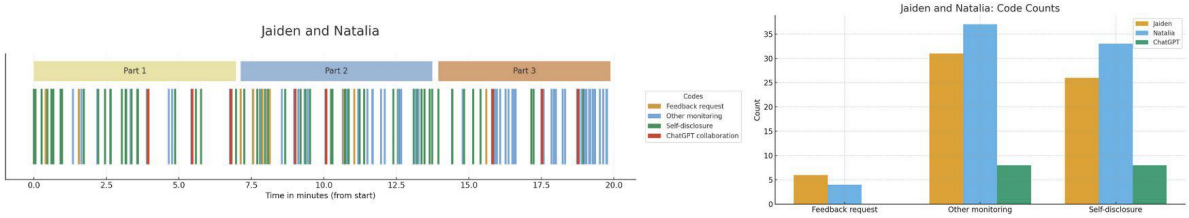
Jaiden and Natalia: ChatGPT is a collaborator

Jaiden and Natalia used ChatGPT in ways that went beyond the task's initial goals. In addition to reminding them about algebraic concepts or processes that they had forgotten, it helped them to create a visual of the roller coaster that they later used to verify the correctness of their solution for increasing and decreasing intervals (task part 2). Jaiden and Natalia used ChatGPT for every portion of the task, even to validate ideas (such as domain and range) that they already knew (see Figure 4). While they exhibited similar patterns of other monitoring as Alex and Kam (see Figure 2), they tended to monitor ChatGPT as the "other" and attempt to understand its thinking, whereas Alex and Kam responded to each other's thinking and reasoning. Natalia assumed a role as communicator with

ChatGPT and prompted it nine times throughout the task while Jaiden tended to state her ideas aloud then write them down until Natalia/ChatGPT gave a contrasting idea or solution. For example:

> Jaiden (8:35) (*writes interval*) Zero…to 32...is it union? Or the other one? [feedback request]
> Natalia (9:00) I'm going to ask [ChatGPT]...how do you do it. [other monitoring]
> Jaiden (9:07) I used to be good at this…okay I'm just going to write A to B. [self-disclosure]
> Natalia (9:32) (*reading from ChatGPT*) Oh, it's one of these things… [other monitoring]
> Jaiden (9:47) What?...was it union? [other monitoring]
> Natalia (9:49) Yeah, it was like that. [other monitoring]

Figure 4
Jaiden and Natalia's problem solving with ChatGPT



Collaboration codes and ChatGPT interactions over time (left) and collaboration frequencies (right).

About halfway through the task, Natalia noticed that ChatGPT gave problem solving advice that went beyond the scope of their prompt. For instance, when she asked it "*would you write [interval] notation based on the x value or the y value,*" (10:30) she observed that it returned advice about what to do ("*describe the behavior over x-intervals*") as well as what not to do "⚠️ *Don't do this:* ❌ *Range-style notation like [0,∞) (that's for y-values*)"). Natalia and Jaiden interpreted this advice as humanistic, and wondered what additional capacities ChatGPT might hold. Natalia spent the remainder of the task prompting ChatGPT to make an accurate graph of the roller coaster function, and was impressed when it created a curve with points A-E plotted that looked extremely similar to the task. They appraised the quality of ChatGPT's curve:

> Natalia (18:10): I got it! [other monitoring]
> Jaiden (18:11): That's so cool, technology's crazy. [other monitoring]
> Natalia (18:18): From what I notice though, point B's not at… [other monitoring]
> Jaiden (18:23): …the maximum? [other monitoring]
> Natalia (18:24): Yeah. That's still pretty good though. I didn't think that [ChatGPT] was going to understand what I was trying to say. [other monitoring]

In this interaction, Natalia means that she did not think ChatGPT would be able to create accurate curves. Jaiden elaborated later, saying that "I've asked AI to give me a graph before, and it's so bad, like not even helpful…and this is actually like, whoa, the best I've seen so far." Both Jaiden and Natalia's expansive uses of ChatGPT for nonroutine aspects of problem solving enabled them to frame it as more of a collaborator than a tool, as it was able to "understand" their request to create a graph and do so with higher-than-expected accuracy. While the graph it created did not give them new information or help them solve additional components of the task, it may have expanded their perspective about what ChatGPT was capable of doing.

Connor and Colin: No ChatGPT collaboration

We briefly highlight one student dyad, Connor and Colin, who did not use ChatGPT to solve the task. Instead, they conversed back-and-forth by monitoring each other's ideas and asking for feedback from each other, as shown here as they mused over ideas about the roller caster's range:

> Connor (1:46): Okay technically, the range can do whatever you want. [self-disclosure].
> Colin (1:50): I thought the range was from [point A to E]. Is that not right? [feedback request]
> Connor (1:54): Well what if later in the coaster, it goes infinitely high? [other monitoring]
> Colin (1:56): Oh, that's true! [other monitoring]

 © ISLS

During task part 5, Colin and Connor diverged in how they were thinking about collaborating with ChatGPT. Colin alluded that he would have used ChatGPT to verify the accuracy of their strategy, while Connor's comments hint that ChatGPT would have not given them opportunities to approach the task creatively:

> Connor (17:47): So I don't know if you can just do this... [self-disclosure]
>
> Colin (17:59): Personally, I would have used AI for that one. [self-disclosure]
>
> Connor (18:00): I mean, I was just imagining for this one…it's an interesting, creative assignment. [self-disclosure]
>
> Colin (18:14): It's a low stakes assignment…I don't need to get it completely right. [self-disclosure]

Discussion and conclusions

Our findings reveal that students solved the task with ChatGPT in ways that differentiated their cognitive, social, and metacognitive work (An et al., 2024), and held implications for what they may have learned. Alex and Kam's episode appeared to have similarities to collaborative problem solving before ChatGPT, as they did not use its results to reconfigure their cognitive and social dynamics, and even forgot that they had asked ChatGPT to procure a solution. Jaiden and Natalia, on the other hand, spent considerable time making sense of and monitoring ChatGPT's contributions, and transformed the goal of the task (Hughes, 2005) into investigating what types of visualizations ChatGPT could produce. Elsa and Gretchen fell somewhere in-between, drawing on ChatGPT as a more knowledgeable other during instances where they felt they could not proceed without more information. However, while they maintained social and metacognitive aspects of collaboration, they did not tend to build upon each other's thinking in ways that suggested that they held a deep understanding of the task's algebra concepts (Confrey, 1990). These patterns suggest that when students use LLMs during problem solving, including positioning LLMs in pro-social collaborative roles, they must negotiate whose responses will be taken up and the basis for which mathematical authority is conferred in group settings.

ChatGPT also appeared to influence students' trajectories in engaging in rich mathematical problem solving interactions. Alex and Kam, by monitoring each other's strategies carefully, critiqued each other, jointly revised their ideas, and developed shared meaning about what each question in the task was asking them to do. The other two focal groups tended to draw upon ChatGPT when they were stuck and took up its output before it was clear that they understood what was being asked of them. For instance, Elsa and Gretchen accepted ChatGPT's parabolic equation without verifying that it was correct (it was incorrect, as they had entered points A, B, and C incorrectly) and without confirming that they knew what it meant. This pattern aligns with concerns raised by Bastani et al. (2025), who found that students who collaborated with AI performed worse once it was removed, suggesting potential trade-offs between immediate problem-solving success and longer-term skill development. Our findings extend this concern to multi-human collaboration, where the presence of an LLM may inadvertently reduce the metacognitive monitoring that emerges through peer explanation and negotiation.

Our findings hold implications for how educators integrate LLMs into collaborative learning environments. If groups that treat ChatGPT as a collaborator spend more time monitoring its output than their own ideas, instructors must consider whether such shifts in students' attention undermine the epistemic goals of learning how to think mathematically, collaboratively. Connor's observation that ChatGPT might constrain creative approaches to problem solving suggests that some students already recognize potential costs of collaborating with LLMs, even in low-stakes settings. This raises questions about when teachers should introduce (if at all) LLMs into problem solving when students are in the process of developing expertise, including learning how to build upon one another's thinking. Our study suggests that when students do not have mathematical ideas of their own, or become stuck, they revert to LLMs' cognitive authority and information efficiency rather than co-developing mathematical conjectures together through social negotiation (Goos et al., 2002). As AI systems become integrated into education settings, teachers, designers and researchers must grapple with how to preserve cognitive, social, and metacognitive dimensions of human collaboration when possible, so that students can develop their own epistemic agency independent of the capacities of AI systems.

At the same time, as LLM-based technologies evolve, we anticipate that students' and teachers' notions of how to collaborate with them will shift (e.g., Delaney et al., 2025). One unresolved question is the extent to which students in our study were ever taught differentiated ways to collaborate with LLM-based tools (Cukurova, 2025). We suspect they may not have been taught, since most used ChatGPT as an information-retrieval machine rather than a collaborator for co-constructing ideas. We thus view our work as contributing a baseline for how students choose to involve LLM-based tools as collaborators absent any sort of instructional intervention or targeted recruitment of students who hold particular epistemic beliefs about collaborating with AI. We aim to address instructional interventions for student-student-LLM mathematical collaboration in future research.

References

Adisa, I. O., & Adefisayo, A. O. (2025). Middle school students' perspectives on adopting generative AI in K-12 education. *The Journal of Educational Research*, *118*(6), 724-735.

An, S., Zhang, S., Cai, Z., Pan, W., Li, M., & Tong, M. (2025). Revealing the interplay of cognitive, meta-cognitive, and social processes in university students' collaborative problem solving: A three-stage analytical framework. International Journal of Computer-Supported Collaborative Learning, 20(1), 9-39. https://doi.org/10.1007/s11412-024-09429-0

Andrews-Todd, J., & Forsyth, C. (2020). Exploring social and cognitive dimensions of collaborative problem solving in an open online simulation-based task. *Computers in Human Behavior, 104*, 105759. https://doi.org/10.1016/j.chb.2018.10.025

Barron, B. (2003). When smart groups fail. *The Journal of the Learning Sciences, 12*(3), 307–359. https://doi.org/10.1207/S15327809JLS1203_1

Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. Proceedings of the National Academy of Sciences, 122(26), e2422633122.

Boaler, J. (2016). Mathematical mindsets: Unleashing students' potential through creative math, inspiring messages, and innovative teaching. Jossey-Bass.

Çini, A., Järvelä, S., Dindar, M., & Malmberg, J. (2023). How multiple levels of metacognitive awareness operate in collaborative problem solving. *Metacognition and Learning, 18*(3), 891–922. https://doi.org/10.1007/s11409-023-09358-7

Confrey, J. (1990). What constructivism implies for teaching. In R. B. Davis, C. A. Maher, & N. Noddings (Eds.), *Constructivist views on the teaching and learning of mathematics* (JRME Monograph No. 4, pp. 107–122). National Council of Teachers of Mathematics. https://doi.org/10.2307/749916

Cukurova, M. (2025). The interplay of learning, analytics and artificial intelligence in education: A vision for hybrid intelligence. British Journal of Educational Technology, 56, 469–488. https://doi.org/10.1111/bjet.13514

Davis, N. (2013). Human-computer co-creativity: Blending human and computational creativity. In Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (Vol. 9, No. 6, pp. 9-12). https://doi.org/10.1609/aiide.v9i6.12603

Delaney, V., Adisa, I. O., Mah, C., & Lee, V. R. (2025). Teaching high school students about generative AI: Cases of teacher lesson design. *The Journal of Educational Research*, *118*(6), 736-751. https://doi.org/10.1080/00220671.2025.2510415

Delaney, V., Stein, S., Sawi, L., & Holliday, K. H. (2025, October). Undergraduate Students' Appraisals and Rationales of AI Fairness in Higher Education. In *Proceedings of the Artificial Intelligence in Measurement and Education Conference (AIME-Con): Full Papers* (pp. 329-336). https://aclanthology.org/2025.aimecon-main.35/

Dindar, M., Järvelä, S., & Järvenoja, H. (2020). Interplay of metacognitive experiences and performance in collaborative problem solving. *Computers & Education*, *154*, 103922. https://doi.org/10.1016/j.compedu.2020.103922

diSessa, A. A. (2007). An interactional analysis of clinical interviewing. *Cognition and Instruction, 25*(4), 523–565. https://doi.org/10.1080/07370000701632413

Goos, M., Galbraith, P., & Renshaw, P. (2002). Socially mediated metacognition: Creating collaborative zones of proximal development in small group problem solving. *Educational Studies in Mathematics*, *49*(2), 193–223. https://doi.org/10.1023/A:1016209010120

Hesse, F., Care, E., Buder, J., Sassenberg, K., Griffin, P. (2015). A framework for teachable collaborative problem solving skills. In P. Griffin, & E. Care (Eds.), Assessment and Teaching of 21st Century Skills(pp. 37–56). Educational Assessment in an Information Age. Springer, Dordrecht. https:// doi. org/ 10.1007/ 978-94- 017- 9395-7_2

Hossein-Mohand, H., Hossein-Mohand, H., Albanese, V., & Olmos Gómez, M. D. C. (2025). AI in mathematics education: A bibliometric analysis of global trends and collaborations (2020-2024). Eurasia Journal of Mathematics, Science and Technology Education, 21(2), em2576. https://doi.org/10.29333/ejmste/15915

Huang, A. Y. Q., Lu, O. H. T., & Yang, S. J. H. (2023). Effects of artificial Intelligence–Enabled personalized recommendations on learners' learning engagement, motivation, and outcomes in a flipped classroom. Computers & Education, 194, Article 104684. https://doi.org/10.1016/j.compedu.2022.104684

Hughes, J. E. (2005). The role of teacher knowledge and learning experiences in forming technology-integrated pedagogy. *Journal of Technology and Teacher Education, 13*(2), 277–302.

Kaput, J. J. (1992). Technology and mathematics education. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 515–556). Macmillan.

Kumar, H., Vincentius, J., Jordan, E., & Anderson, A. (2025). Human creativity in the age of LLMs: Randomized experiments on divergent and convergent thinking. In Proceedings of the 2025 CHI Conference on human Factors in computing systems. Yokohama Japan: CH|'25: CHl Conference on Human Factors in Computing Systems. https://doi.org/10.1145/3706598.3714198.

Jordan, B., & Henderson, A. (1995). Interaction analysis: Foundations and practice. *The Journal of the Learning Sciences, 4*(1), 39–103. https://doi.org/10.1207/s15327809jls0401_2

Li, S., Pöysä-Tarhonen, J., & Häkkinen, P. (2022). Patterns of action transitions in online collaborative problemsolving: A network analysis approach. International Journal of Computer-supported Collaborative Learning, 17(2), 191tional. https://doi. org/10.1007/s11412-022-09369-7

Liu, J., Sun, D., Sun, J., Wang, J., & Philip, L. H. Y. (2025). Designing a generative AI enabled learning environment for mathematics word problem solving in primary schools: Learning performance, attitudes and interaction. Computers and Education: Artificial Intelligence, 100438.

Mamykina, L., Candy, L., & Edmonds, E. (2002). Collaborative creativity. *Communications of the ACM, 45*(10), 96–99. https://doi.org/10.1145/570907.570940

OECD. (2017). PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematic, *Financial Literacy and Collaborative Problem Solving*, revised edition, PISA, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264281820-en

Ouyang, F., Xu, W., & Cukurova, M. (2023). An artificial intelligence-driven learning analytics method to examine the collaborative problem-solving process from the complex adaptive systems perspective. International Journal of Computer-Supported Collaborative Learning, 18(1), 39–66. https://doi.org/10.1007/s11412-023-09387-z

Remillard, J. T. (2005). Examining key concepts in research on teachers' use of mathematics curricula. *Review of educational research, 75*(2), 211-246. https://doi.org/10.3102/003465430750022

Rezwana, J., & Maher, M. L. (2021). COFI: A framework for modeling interaction in human-AI co-creative systems. *Proceedings of the Second Workshop on the Future of Co-Creative Systems (ICCC 2021)*. https://doi.org/10.1145/351902

Rizos, I., & Gkrekas, N. (2025). The impact of LLMs on mathematics education and research at the university. *Social Sciences & Humanities Open, 12*, 101969. https://doi.org/10.1016/j.ssaho.2025.101969

Roschelle, J., Teasley, S.D. (1995). The Construction of Shared Knowledge in Collaborative Problem Solving. In: O'Malley, C. (eds) Computer Supported Collaborative Learning. NATO ASI Series, vol 128. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-85098-1_5

Scardamalia, M., & Bereiter, C. (2020). Will knowledge building remain uniquely human? *Qwerty – Open and Interdisciplinary Journal of Technology, Culture and Education, 15*(2), 12–26. https://doi.org/10.30557/QW000028

Slavin, R. E. (2014). Cooperative learning and academic achievement: Why does groupwork work? *Anales de Psicología, 30*(3), 785–791. https://doi.org/10.6018/analesps.30.3.201201

Subramonyam, H., Pea, R., Pondoc, C., Agrawala, M., & Seifert, C. (2024, May). Bridging the gulf of envisioning: Cognitive challenges in prompt based interactions with LLMs. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (pp. 1-19).

Sun, C., Shute, V. J., Stewart, A., Yonehiro, J., Duran, N. D., & D'Mello, S. K. (2020). Towards a generalized competency model of collaborative problem solving. Computers & Education, 143, 103672. https://doi.org/ 10.1016/j.compe du.2019.103672

Swiecki, Z., Ruis, A. R., Farrell, C., & Shaffer, D. W. (2020). Assessing individual contributions to collaborative problem solving: A network analysis approach. *Computers in Human Behavior, 104*, 105876. https://doi.org/10.1016/j.chb.2019.01.009

Urmeneta, A., & Romero, M. (Eds.). (2024). Creative applications of artificial intelligence in education. Nature Switzerland: Springer. https://doi.org/10.1007/978-3-031- 55272-4.

Yin, R. K. (2017). *Case study research and applications: Design and methods* (6th ed.). SAGE Publications.

International Society of the Learning Sciences

Quantifying Knowledge Convergence through Change Point Detection of Word Abstractness in Collaborative Discussions

Ryunosuke Nishimura, Hironori Egi
r.nishimura@uec.ac.jp, hiro.egi@uec.ac.jp
The University of Electro-Communications, Japan

Abstract: The quantification of knowledge convergence often relies on qualitative coding, limiting insights into how the dispersion of abstractness evolves during collaborative discussion. This study proposes a quantitative method that models each discussion as a time series of abstractness and applies change point detection to identify distributional shifts in its mean and dispersion. Six groups engaged in ten-minute discussions. Nouns in their utterances were extracted and assigned abstractness scores, after which change-point detection was applied to each group's abstractness sequence with a minimum segment size of 20%, and the number of change points was determined by BIC. Results showed that the mean followed a U-shaped trajectory while the standard deviation showed the inverse in four groups, reflecting a mid-discussion phase in which abstract reasoning was anchored in concrete examples. These results demonstrate that knowledge convergence appears as interpretable shifts thereby bridging qualitative analysis of knowledge convergence with quantitative segmentation.

Introduction

Research on collaborative discussion has highlighted how individuals integrate ideas distributed among participants with their prior knowledge, thereby deepening understanding and generating new insights into a shared meaning (Sawyer, 2014; Stahl, 2002). In CSCL contexts, collaborative discussions are incorporated in diverse forms, including online scaffolding for collaboration (Zheng et al., 2022) and holographic agents that promote group regulation (Nishimura et al., 2023). Analyses of collaborative discussions are often grounded in the framework of knowledge convergence, which examines how a group's understandings become shared over time (Jeong & Chi, 2007; Weinberger et al., 2007). Knowledge convergence involves recurrent movement between concrete and abstract concepts; learners construct shared knowledge by abstracting from episodes and by returning to examples to test and refine those abstractions (Roschelle, 1992; Schwartz, 1999). Studying collaborative discussions through the lens of knowledge convergence reveals how learners transform individual understandings into shared knowledge by abstracting from concrete experiences and jointly refining those abstractions (Dillenbourg, 1999; Jorczak, 2011). However, there remains a widely recognized need to quantify such processes in ways that connect theories of learning with computational approaches to discussion analysis (Jeong et al, 2014, Wise & Schwarz, 2017). Prior work often relies on qualitative analyses that categorize utterances into a small set of predefined states, such as concrete or abstract (Wilmont et al., 2012). Therefore, there is a lack of insight into how to determine appropriate state counts and how to systematically capture the dynamic widening or narrowing of abstractness in the discussion, rather than merely tracking directional changes between concrete and abstract states.

This study introduces a quantitative approach to analyzing knowledge convergence by applying change point detection to lexical abstractness sequences, building on prior work that quantified the convergence based on the abstractness of words in utterances (Nishimura & Egi, 2025). In this study, knowledge convergence is operationalized at the level of group discourse as temporal shifts in the abstractness of words used in the discussion, rather than as a measure of whether individual participants have reached identical internal understanding. Our objective is to provide quantitative insights into knowledge convergence, focusing on capturing dynamic changes in both the mean and dispersion of the abstractness throughout the discussion. As an initial validity check, we examine how the data-driven patterns align with the actual content of group discussions, thereby validating of the quantitative insights into knowledge convergence.

Literature review

Knowledge convergence in collaboration

Knowledge convergence refers to the gradual integration of shared knowledge in interaction, a process that oscillates between concrete experiences and abstract formulations (Roschelle, 1992). Building on this definition,

 © ISLS

research on knowledge convergence has developed along two complementary views: a process-oriented view that analyzes how convergence unfolds during interaction and an outcome-oriented view that examines convergence in post-hoc learning products (Fischer & Mandl, 2005).

From a process-oriented view, researchers examine how participants adjust the level of the abstractness in their utterances, shifting between concrete experiences and abstract concepts during interaction (Weinberger et al., 2007). Accordingly, knowledge convergence is understood as involving both abstraction and concretion, rather than focusing on one over the other. More specifically, prior work characterizes knowledge convergence as an iterative, back-and-forth movement between concrete and abstract ideas, including both the generalization from concrete examples and the re-grounding of abstract knowledge through concrete instances (Wilmont, 2012; Chaparro-Moreno et al., 2023). Consistent with this view, abstract external representations such as concept maps and argument diagrams have been shown to guide collaborative knowledge convergence (Janssen et al., 2010; Kolloffel et al., 2011; Van Boxtel et al., 2000). In addition, concrete external representations such as heuristic worked examples also promote abstraction and thereby foster knowledge convergence (Kollar et al., 2014). However, high-quality processes do not guarantee high-quality outcomes: groups may display sophisticated abstraction and still fail to share understanding across members (Barron, 2003). Related perspectives also say that the process does not ensure learning outcomes (Kreijns et al., 2003; Kirschner et al., 2009). While the importance of abstractness in the process of knowledge convergence has been recognized, researchers hold divergent views on its influence on learning outcomes. This uncertainty partly stems from the fact that most previous studies have relied on qualitative coding, without a consistent understanding of what is considered concrete or abstract. To address this issue, it is essential to establish a quantitative measure of abstractness applicable across discussions. In particular, quantifying abstractness as a continuous variable allows for capturing its mean level and dispersion over time, thereby complementing insights obtained through qualitative research.

To summarize, although concretion and abstraction are known to occur throughout the process of knowledge convergence, quantitative insights into the magnitude and dynamics of shifts in abstractness remain limited. This lack of quantitative understanding makes it difficult to describe how abstractness changes and develops during collaborative interaction. Therefore, quantifying the abstractness of utterances within the flow of interaction can provide a clearer view of how learners construct and refine shared understanding during collaboration.

Analysis of knowledge convergence

Recent advances in information technology have broadened the scope of quantitative analyses of how learners engage in collaborative discussions. Within this research domain, studies have primarily focused on utterances and observable actions rather than on learners' internal understanding processes. This focus arises because utterances and actions are directly observable and can be readily converted into digital data suitable for computational analysis. Although their interpretation may differ across individuals, these behaviors are consistently measurable and therefore well suited for quantitative investigation. In contrast, understanding during discussion is highly individualized, and spoken expressions do not necessarily align with participants' internal comprehension (Shirouzu & Miyake, 2002; Cang et al., 2024). Therefore, most quantitative approaches to the knowledge convergence have tended to focus on externalized knowledge (Suthers et al., 2007; Dehler et al., 2011; Zheng et al., 2014) or observable actions (Zottmann et al., 2013; Xiang et al., 2024).

Among quantitative studies on knowledge convergence, research on abstraction typically employs qualitative coding of utterance abstractness and analyzes transitions across abstractness levels to examine how discussions shift between concrete and abstract modes of expression. For example, utterances in idea-generation sessions were coded into two states, concrete and abstract, and their temporal transitions were examined (Antoniou et al., 2019). The results indicated that concrete utterances dominated during divergent phases, whereas both concrete and abstract utterances alternated during convergent phases. Subsequent research proposed a three-level coding framework comprising concrete, first-order abstraction, and second-order abstraction (Wilmont et al., 2020). In this framework, first-order abstraction refers to identifying shared patterns across concrete examples, whereas second-order abstraction involves deriving higher-order principles underlying those patterns. Their analysis suggested that knowledge convergence tends to progress from concrete examples to first-order and subsequently to second-order abstractions; however, this process is cyclical rather than linear, characterized by recurring shifts between abstraction and concretization. Using the four-level model of abstraction as a framework (Blank et al., 1978), later research likewise demonstrated bidirectional transitions among abstraction levels in preschool conversations (Chaparro-Moreno et al., 2023).

Although these studies have established detailed frameworks for coding abstraction levels, three major limitations remain. First, the quantification of abstractness has not been automated, which makes it difficult to conduct cross-group comparisons or to quantify discussions in real time. Second, because the unit of analysis is

confined to individual utterances, these approaches cannot capture the abstractness dynamics that unfold across multiple utterances. Third, while these studies describe hierarchical levels of the abstractness, they do not examine how the abstractness becomes more concentrated or more dispersed as the discussion progress. This distinction matters because knowledge convergence is theorized as repeated movement between concrete grounding and abstraction; a discussion phase may differ not only in its average level of the abstractness but also in whether concrete and abstract expressions co-occur within it. In short, previous studies have measured how high or low the level of the abstractness is, but not how wide or narrow its range becomes. To address the first two of these limitations, our previous research introduced methods for visualizing the abstractness of words in utterances and quantifying collaborative discussions using the moving average of the abstractness (Nishimura et al., 2024; Nishimura & Egi, 2025). These methods made it possible to observe gradual shifts in the abstractness over time. However, the analyses primarily focused on mean-level trends. The dynamic concentration and dispersion of the abstractness, reflecting its dispersion, have not yet been systematically investigated.

In summary, previous studies that quantified the abstractness in knowledge convergence have mainly focused on its mean level, overlooking how it becomes more concentrated or more dispersed over time. Thus, a new quantitative framework is needed to capture both the level and variability of abstractness and reveal a finer structure of knowledge convergence.

Research questions

In summary, previous research has provided valuable insights into knowledge convergence by quantifying the abstractness of utterances, yet these approaches have primarily focused on mean levels of abstractness and overlooked its temporal dispersion. To address this gap, we propose the method to quantify both the average and variability of the abstractness to capture how the abstractness changes during knowledge convergence. By analyzing change point of the abstractness, we aim to characterize the dynamic structure of knowledge convergence with greater precision. The research question is: “How are the dynamic features of knowledge convergence reflected in the temporal changes of both the mean and dispersion of abstractness in collaborative discussions?”

Methods

Quantifying the abstractness in participants utterances

Since the participants were Japanese, the abstractness of words in their utterances was quantified using the AWD-J EX dictionary (Nara Institute of Science and Technology, 2019), a large-scale lexical resource containing approximately 437,000 Japanese words. The AWD-J EX was designed for research on cognitive processing and assigns every word an abstractness rating from 1 to 5, where smaller values represent more concrete meanings and larger values indicate higher abstractness. For instance, in this lexicon, Tokyo has a rating of 1.3, city 2.0, and optimization 2.5. Notably, words with abstractness ratings above 3 often include nouns, adjectives, and verbs that are not commonly used in everyday language. Each utterance was thus converted into a numerical sequence representing the abstractness of the words it contained.

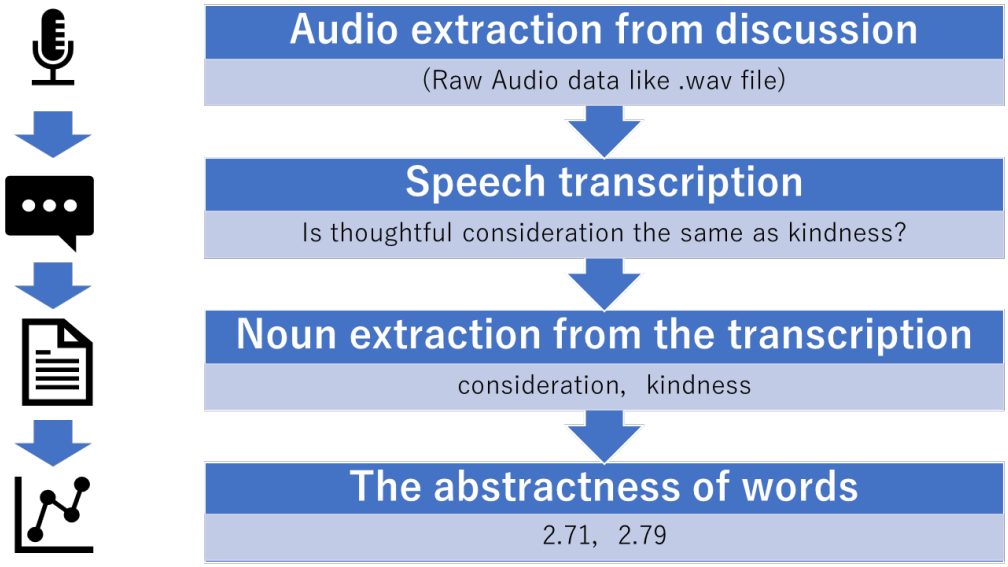
Discussion dataset and preprocessing

The discussion corpus used in this study was drawn from sessions conducted at a Japanese university with undergraduate and graduate students majoring in science and engineering. Each of the six groups, consisting of three participants, engaged in a ten-minute collaborative discussion on the question “What is happiness?”. This topic was selected because it was familiar to all participants and encouraged them to share concrete experiences and express their ideas in their own words (Nishimura et al., 2023). To facilitate knowledge convergence, all groups were instructed to reach a shared conclusion by the end of the discussion.

Figure 1 provides an overview of the data processing pipeline used in this study. Each participant’s speech was recorded using an individual microphone connected to a single-board computer. The resulting audio files were automatically transcribed with the Whisper API (OpenAI, 2024) and were subsequently reviewed manually to correct recognition errors. After transcription, morphological analysis was applied to the text to identify parts of speech. From this analysis, nouns were extracted as the primary target for abstractness measurement because nouns most directly index the conceptual objects and provide a relatively stable unit for tracing topic-level shifts in the abstractness across utterances. Each noun was then assigned an abstractness using the AWD-J EX dictionary, which rates words on a scale from 1 (concrete) to 5 (abstract). The resulting values were arranged in order of occurrence to form a numerical sequence for each utterance (e.g., [2.0, 2.2, 2.4, 2.1, 2.7, …]). These sequences provided the foundation for the subsequent quantitative analyses.

 © ISLS

Figure 1
Data Processing Pipeline to Quantify the Abstractness of Words in Utterance



Change point detection of the abstractness

To capture changes in both the mean and dispersion of the abstractness during discussions, we applied change point detection to numerical sequences representing abstractness. Change point detection is a statistical method that identifies points in a time series where the underlying data distribution, such as the mean or variance, changes significantly (Aminikhanghahi & Cook, 2016). This method has been applied in a variety of research fields, including finance, climate science, and biomedical signal analysis, to detect transitions in underlying processes or behavioral patterns (Truong et al, 2019). Its flexibility and reliability make it suitable for identifying transitions in temporal data, such as changes in the abstractness during collaborative discussions.

In this study, we used the Dynp method implemented in the Python library ruptures (Truong et al., 2019). The library provides modular implementations of several algorithms for segmenting time-series data. The Dynp method requires two parameters to be specified: the *model* parameter, which defines the assumed statistical distribution of the data, and the *min_size* parameter, which sets the minimum interval length between adjacent change points to prevent over-segmentation caused by small fluctuations or noise. In our analysis, the *model* parameter was set to a normal distribution because the abstractness values were continuous and unimodal, which made this distribution appropriate for capturing gradual variations in the data. The normal model is commonly adopted in change point detection for continuous time-series data (Aminikhanghahi and Cook, 2016; Truong et al., 2019). The *min_size* parameter was empirically set to 20% of the total number of nouns in each discussion. This setting was chosen to avoid detecting change points within very short segments such as individual utterances, which would make interpretation unreliable. Because the total number of nouns varied among discussions, the value of *min_size* was adjusted for each dataset. To assess the robustness of this empirical setting, we repeated the analysis with *min_size* values varied by ±10% and examined whether the number of detected segments remained consistent. The Dynp algorithm also requires specifying the number of change points, denoted as *K*, in advance. To determine the optimal segmentation, we tested a range of *K* values, starting from zero and increasing up to the theoretical upper limit, which corresponds to the total number of data points divided by *min_size*, minus one. The optimal number of change points was selected according to the Bayesian Information Criterion (BIC).

As a result, each abstractness sequence was divided into several segments that differed in their means or variances. Comparing these segments allowed us to examine how discussions evolved over time, for example, from concrete to abstract expressions or from narrower to broader ranges of the abstractness.

Results

Figure 2 shows the sequential data of the abstractness and change points for all groups. The vertical axis represents the degree of abstractness, and the horizontal axis indicates the order in which nouns appeared. The vertical dotted lines denote the change points detected in each sequence. Table 1 presents, for each interval divided by the detected change points, the number of nouns and the corresponding mean and SD of the abstractness. In the change

point detection procedure, the minimum segment size was set to 20% of the total number of nouns in each group. The total numbers of nouns were 319, 343, 327, 368, 312, and 327 for each Group, resulting in minimum segment sizes of 63, 68, 65, 73, 62, and 65, respectively. A sensitivity analysis varying this parameter within ±10% of its default value showed that the number of detected change points remained unchanged for all groups except Group 1; in Group 1, the number of detected change points varied between two and three. This indicates that the results were largely robust to parameter variation.

Figure 2

Change Point of the Abstractness for Each Group

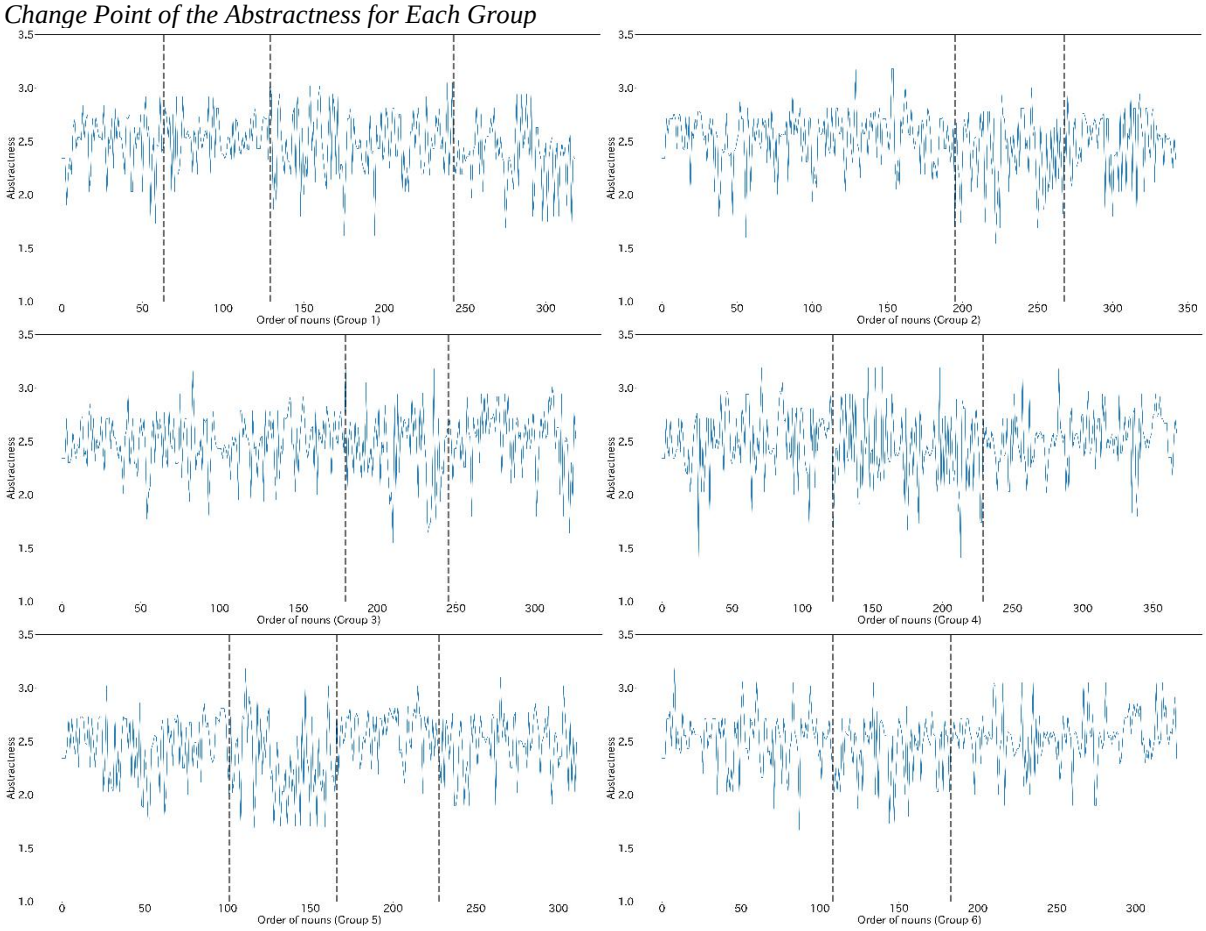


Table 1

Segment-Wise Means and Dispersions of Abstractness Defined by the Detected Change Points.

	Segment	The number of words	Mean Abstractness	SD of Abstractness
Group 1	[1, 63]	63	2.473	0.268
	[64, 129]	66	2.534	0.206
	[130, 243]	114	2.475	0.313
	[244, 319]	76	2.365	0.289
Group 2	[1, 195]	195	2.532	0.247
	[196, 268]	73	2.368	0.348
	[269, 343]	75	2.499	0.276
Group 3	[1, 180]	180	2.486	0.240
	[181, 245]	65	2.418	0.367
	[246, 327]	82	2.540	0.292
Group 4	[1, 122]	122	2.509	0.293
	[123, 229]	107	2.405	0.365
	[230, 368]	139	2.521	0.240
Group 5	[1, 101]	101	2.454	0.282
	[102, 166]	65	2.288	0.381

 © ISLS

	[167, 228]	62	2.584	0.208
	[229, 312]	84	2.462	0.273
Group 6	[1, 108]	108	2.522	0.254
	[109, 183]	75	2.412	0.289
	[184, 327]	144	2.540	0.238

Discussion

Change point of the abstractness of words in utterances

Knowledge convergence is commonly characterized as a bidirectional movement between concretion and abstraction through which participants align and integrate their understandings (Roschelle, 1992; Wilmont et al., 2012). In this study, the process of knowledge convergence is quantified as a time series of the abstractness and change-point detection is applied to identify moments at which the underlying mean and/or dispersion of abstractness shifted. Based on the method, we examined how the dynamic features of knowledge convergence are reflected in the temporal changes in the abstractness.

Figure 2 shows that the sequences are jagged, with frequent short-range oscillations. This profile is not merely noise. It reflects the way speakers illustrate general ideas through concrete examples: one might say they feel “happiness” when they read a “book” or drink a “coffee”. This pattern aligns with previous findings suggesting that abstract concepts are often anchored in internal states and tend to elicit utterances drawing on concrete or experiential content (Villani et al., 2022). Figure 2 also provides the basis of the first quantitative result, concerning the number of segments inferred from the detected change points. The optimal segmentation differed across groups: Groups 2, 3, 4, and 6 were divided into three segments, whereas Groups 1 and 5 were divided into four segments. For each discussion, the minimum segment length was set to 20% of the total number of words, allowing at most one long segments in theory. Yet, none of the series collapsed to two or fewer segments. Thus, every group showed at least two shifts in the mean or dispersion of the abstractness. This pattern reflects the iterative nature of knowledge convergence, where participants alternate between examples and generalizations rather than moving steadily in one direction. The variation in segment count across groups further suggests that analytic interpretations should account for group-specific phase structures instead of assuming a single uniform progression.

In addition, Table 1 shows the segment-wise tendencies of abstractness across intervals defined by the detected change points. Across the four groups divided into three segments (Groups 2, 3, 4, and 6), the mean of the abstractness followed a U-shaped trajectory, declining from the first to the second segment and rising again in the third, while the standard deviation showed the inverse pattern. This pattern reflects a mid-discussion state in which utterances were anchored in concrete examples yet bridged to abstract ideas, lowering the mean of the abstractness while increasing its variability. To clarify how the U-shaped tendency appeared in the actual discussions, the following analysis examines representative utterances from each group. Quoted nouns are annotated with an abstractness value in parentheses. Because lexical segmentation differs between Japanese and English, the number and boundaries of nouns in quotations do not always align across the two languages. For example, an English expression such as “time limit” may correspond to a single Japanese noun. With this caveat in mind, the following examples illustrate how participants alternated between concrete and abstract references.

Group 3 initially discussed two broad categories of “happiness” (2.71): “material” (2.16) and “mental” (2.78), while also mentioning “volunteer” (1.95) as an illustrative example. In the middle segment, participants referred to the fact that one’s “stomach” (2.02) becomes empty in the “evening” (1.55), using this concrete experience to explore whether happiness has a “time limit” (2.21). In the final segment, they reached agreement that “material” (2.16) happiness is “temporary” (2.73) whereas “mental” (2.78) happiness is “lasting” (2.93), concluding that happiness is the “resolution” (2.57) of “desires” (2.94). Group 6 followed a similar course. The discussion began with shared understanding that happiness is “subjective” (2.68) and “relative” (2.59). In the middle part, the participants considered an “elderly person” (1.81) recalling the “past” (2.56) and examined how reinterpretation or idealization of the past might alter the feeling of happiness. By the end, they organized their view around the two interrelated dimensions of “value” (2.52) and “time” (2.56), proposing that happiness is shaped by how people assign value to the past and present. Group 2 also began with an abstract discussion on how the duration of “time” (2.56) spent under “difficulty” (2.57) relates to happiness. In the middle segment, the group anchored its reasoning in Maslow’s “pyramid” (1.86) of needs (Maslow, 1943) and examined the minimal “condition” (2.65) required for happiness. Finally, they converged on the view that happiness corresponds to a “state” (2.76) in which “life” (2.44) is materially secured and stable. Group 4 started from the fulfillment of “desires” (2.94) such as the satisfaction of “appetite” (2.38). In the middle phase, participants drew on “Instagram” (2.03) “comparison” (2.39) to discuss desires beyond the basic physiological ones. By the end, they concluded

 © ISLS

that happiness requires not only the fulfillment of fundamental needs but also satisfaction with one's "status" (2.57) and overall "state" (2.76).

In contrast, Groups 1 and 5 showed four segments rather than three. Group 5 largely followed the U-shape pattern for its first three segments but exhibited an additional downturn in the final phase, reflecting a late re-grounding of abstract ideas in concrete considerations. At the outset, participants proposed that "happiness" (2.71) consists in the absence of "stress" (2.49), defining it as the opposite of "unhappiness" (2.73). In the middle phase, the discussion turned to more concrete examples, using "club activity" (2.02) as a reference point to compare the degree of "satisfaction" (2.73) when the "result" (2.47) of a competition was favorable or not. In the third segment, participants abstracted these situational examples into a temporal framework, comparing happiness experienced in the "present" (2.54) with that anticipated in the "future" (2.76), and concluded that being satisfied at this very moment constitutes happiness. Finally, the discussion returned to the example of "club activity" (2.02), where participants reflected on their earlier points. Thus, the final phase of Group 5 represented a reconfirmation of its conclusion, consistent with the theoretical notion of knowledge convergence. Group 1, by contrast, showed an inverted trajectory. It reached its conclusion relatively early. In the final phase, participants exchanged remarks unrelated to the discussion itself, such as checking the remaining time (two" (1.8) "minutes" (2.36) left?). Taken together, these characteristics suggest that the interpretive weight of the segments should be considered with caution.

These examples indicate that the processes of concretization and abstraction during knowledge convergence, along with the dispersion and concentration of abstractness, can be quantitatively captured as segment-wise shifts in the mean and dispersion of abstractness identified through change-point detection. In our data, four groups exhibited a common structure characterized by a U-shaped trajectory in the mean and an inverted U-shape in the dispersion. The mid-segment, marked by lower mean and higher standard deviation, corresponded to a phase of knowledge convergence in which participants anchored their reasoning in concrete examples while collaboratively integrating more abstract conceptualizations. This pattern is consistent with prior results that knowledge convergence proceeds through iterative movement between concrete grounding and abstraction rather than a linear shift in a single direction (Roschelle, 1992; Wilmont et al., 2020; Chaparro-Moreno et al., 2023). Group 5 added a late re-grounding phase, whereas Group 1 included several remarks unrelated to the discussion. Thus, in response to the research question, the knowledge convergence process of converging abstract ideas through concrete anchoring is reflected as a decrease in the mean and an increase in the standard deviation of abstractness within the segmental structure defined by detected change points.

Limitations

This study has some methodological limitations. Because the analysis dealt with only one discussion topic, the observed patterns may not fully extend to other domains of discourse, particularly those characterized by more concrete reasoning processes. Another limitation involves the fixed parameter settings used in change point detection. Although the robustness of the results was verified within a ±10% range of the *min_size* parameter, examining a wider range of parameter values or alternative segmentation criteria could further clarify the stability of the detected transitions. In addition, although AWD-J EX contains ratings for many words, the present analysis was restricted to nouns; including verbs and adjectives may yield somewhat different patterns of abstractness. The representative excerpts were used as an initial interpretive check on the detected segments, but the study did not include an independent human-coding reliability analysis. Therefore, the present results should be read as a proof of concept rather than a full validation against human coding. The abstractness is also language-dependent to some extent, since the lexical units used for annotation differ between Japanese and English. This means that an English expression may correspond to a single Japanese noun, or vice versa, potentially introducing small cross-linguistic differences in the computed abstractness. Future research should also include multiple discussion topics to evaluate the generality of the present findings across different collaborative contexts.

Conclusion

We proposed a quantitative framework for analyzing the process of knowledge convergence through temporal changes in the abstractness of words in collaborative discussions. Specifically, we modeled each discussion as a time series of abstractness and applied change point detection to identify shifts in the mean and dispersion of the distribution. Data were obtained from six triads who engaged in ten-minute collaborative discussions designed to elicit knowledge convergence. The analysis revealed that knowledge convergence unfolds through alternating phases of concretion and abstraction, observable as distributional shifts in the abstractness. In particular, the process of converging abstract ideas through concrete anchoring was reflected as a decrease in the mean and an increase in the dispersion of the abstractness within segments defined by detected change points. This pattern indicates that participants temporarily grounded abstract reasoning in concrete examples before re-integrating

them into shared abstractions. The segmentation relied on fixed parameters, such as the minimum segment length set to one-fifth of the total nouns. Although the robustness of the results was confirmed through sensitivity checks, adjusting these settings or applying adaptive criteria could enable finer-grained analysis and broaden the interpretability of the detected transitions.

Overall, our method reveals discussion intervals characterized by a wide range of abstractness, that is, intervals where abstract reasoning is anchored in concrete examples. Because the method requires only transcript data and lexical ratings, it can be used to compare phase structures across groups at relatively low annotation cost and to provide quantitative evidence for qualitative analysis. By jointly examining both the mean and dispersion of the abstractness, the approach shows patterns that would remain hidden in analyses focusing on either dimension alone. This dual focus not only extends quantitative frameworks of knowledge convergence but also deepens its qualitative interpretation, providing an integrated analytic perspective on knowledge convergence in collaborative discussions.

References

Aminikhanghahi, S., and Cook, D. J. (2016). A survey of methods for time series change point detection. *Knowledge and Information Systems*, 51(2), 339–367.

Antoniou, R., Dekoninck, E., & Bonvoisin, J. (2019). Exploring the role of linguistic abstraction in idea-generation sessions. *Proceedings of the Design Society: International Conference on Engineering Design* (Vol. 1, No. 1, pp. 329-338). Cambridge University Press.

Barron, B. (2003). When smart groups fail. *The journal of the learning sciences*, *12*(3), 307-359.

Blank, M., Rose, S. A., & Berlin, L. J. (1978). The Language of Learning: The Preschool Years. Grune & Stratton.

Cang, X., Zeng, L., Tissenbaum, M., Gnesdilow, D., & Puntambekar, S. (2024). Convergent Group Understanding: Indicator of Knowledge Integration? In Clarke-Midura, J., Kollar, I., Gu, X., & D'Angelo, C. (Eds.), *Proceedings of the 17th International Conference on Computer-Supported Collaborative Learning* (pp. 197-200). International Society of the Learning Sciences.

Chaparro-Moreno, L. J., Lin, T. J., Justice, L. M., Mills, A. K., & Uanhoro, J. O. (2023). The influence of context on the abstraction level of children's conversations in the preschool classroom. *Early Education and Development*, *34*(3), 705-724.

Dehler, J., Bodemer, D., Buder, J., & Hesse, F. W. (2011). Guiding knowledge communication in CSCL via group knowledge awareness. Computers in Human Behavior, 27, 1068–1078

Dillenbourg, P. (1999). *Collaborative learning: Cognitive and computational approaches. advances in learning and instruction series*. Elsevier Science, Inc., PO Box 945, Madison Square Station, New York, NY 10160-0757.

Fischer, F., & Mandl, H. (2005). Knowledge convergence in computer-supported collaborative learning: The role of external representation tools. *The Journal of the Learning Sciences, 14*(3), 405-441.

Heritage, J., & Watson, D. R. (1979). Formulations as conversational objects. *Everyday language: Studies in ethnomethodology* (pp. 123-162).

Janssen, J., Erkens, G., Kirschner, P. A., & Kanselaar, G. (2010). Effects of representational guidance during computer-supported collaborative learning. *Instructional Science, 38*, 59–88.

Jeong, H., & Chi, M. T. (2007). Knowledge convergence and collaborative learning. *Instructional Science, 35*, 287-315.

Jeong, H., Hmelo-Silver, C. E., & Yu, Y. (2014). An examination of CSCL methodological practices and the influence of theoretical frameworks 2005–2009. *International Journal of Computer-Supported Collaborative Learning, 9*, 305-334.

Jorczak, R. L. (2011). An information processing perspective on divergence and convergence in collaborative learning. *International Journal of Computer-Supported Collaborative Learning*, *6*, 207-221.

Kirschner, F., Paas, F., & Kirschner, P. A. (2009). A cognitive load approach to collaborative learning: United brains for complex tasks. *Educational Psychology Review, 21*, 31–42.

Kollar, I., Ufer, S., Reichersdorfer, E., Vogel, F., Fischer, F., & Reiss, K. (2014). Effects of collaboration scripts and heuristic worked examples on the acquisition of mathematical argumentation skills of teacher students with different levels of prior achievement. *Learning and Instruction, 32*, 22–36.

Kolloffel, B., Eysink, T. H. S., & de Jong, T. (2011). Comparing the effects of representational tools in collaborative and individual inquiry learning. *International Journal of Computer-Supported Collaborative Learning, 6*, 223–251.

Kreijns, K., Kirschner, P. A., & Jochems, W. (2003). Identifying the pitfalls for social interaction in computer-supported collaborative learning environments: *A review of the research. Computers in Human Behavior, 19*(3), 335–353.

Maslow, A. H. (1943). A theory of human motivation. *Psychological Review google schola, 2*, 21-28.

Nara Institute of Science and Technology. (2019). *AWD-J: Abstractness of Word Database for Japanese common words.* https://sociocom.naist.jp/awd-j/

Nishimura, R., Iharada, R., Sugamoto, Y., Ishii, Y., Mochizuki, T., & Egi, H. (2023). Discussion support agent system to promote equalization of speech among participants. *Proceedings of the 31st International Conference on Computers in Education* (Vol 1, pp. 266-271). Asia-Pacific Society for Computers in Education. https://eds.let.media.kyoto-u.ac.jp/ICCE2023/wp-content/uploads/2023/12/ICCE2023-Proceedings-V1-1214-final.pdf#page=329

Nishimura, R., Watanabe, C., & Egi, H. (2024). AbstNebula: A method to visualize the back-and-forth between concrete and abstract utterances for discussion analysis. *Companion Proceedings of the 14th International Conference on Learning Analytics and Knowledge* (pp. 91–93). Society for Learning Analytics Research. https://www.solaresearch.org/wp-content/uploads/2024/03/LAK24_CompanionProceedings.pdf#page=113

Nishimura, R., & Egi, H. (2025). Quantitative Evaluation of Collaborative Discussions Based on the Abstractness of Words in Utterance. In *Proceedings of the 18th International Conference on Computer-Supported Collaborative Learning-CSCL 2025, pp. 12-19*. International Society of the Learning Sciences.

OpenAI. (2024). *Speech to text*. https://platform.openai.com/docs/guides/speech-to-text

Roschelle, J. (1992). Learning by collaborating: Convergent conceptual change. *The Journal of the Learning Sciences*, *2*(3), 235–276.

Sawyer, R. K. (Ed.). (2014). *The Cambridge Handbook of the Learning Sciences* (2nd ed.). Cambridge: Cambridge University Press.

Schwartz, D. L. (1999). *The productive agency that drives collaborative learning*. In P. Dillenbourg (Ed.), *Collaborative learning: Cognitive and computational approaches* (pp. 197–218). New York, NY: Elsevier Science.

Stahl, G. (2002). Contributions to a theoretical framework for CSCL. In G. Stahl (Ed.), *Computer support for collaborative learning: Foundations for a CSCL community* (pp. 62–71). Boulder, CO: International Society of the Learning Sciences.

Shirouzu, H., & Miyake, N. (2002). Learning by collaborating revisited: Individualistic vs. convergent understanding. *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 24, No. 24).

Suthers, D., Vatrapu, R., Medina, R., Joseph, S., & Dwyer, a. (2007). Conceptual Representations Enhance Knowledge Construction in Asynchronous Collaboration. In Chinn, C. A., Erkens, G., & Puntambekar, S. (Eds.), *The Computer Supported Collaborative Learning Conference 2007*, (Vol 8, Part 2, pp. 704-713). International Society of the Learning Sciences.

Truong, C., Oudre, L., & Vayatis, N. (2020). Selective review of offline change point detection methods. *Signal Processing*, *167*, 107299.

Van Boxtel, C., Van der Linden, J., & Kanselaar, G. (2000). Collaborative learning tasks and the elaboration of conceptual knowledge. *Learning and instruction*, *10*(4), 311-330.

Villani, C., Orsoni, M., Lugli, L., Benassi, M., & Borghi, A. M. (2022). Abstract and concrete concepts in conversation. *Scientific Reports*, *12*(1), 17572.

Weinberger, A., Stegmann, K., & Fischer, F. (2007). Knowledge convergence in collaborative learning: Concepts and assessment. *Learning and instruction*, *17*(4), 416-426.

Wilmont, I., Barendsen, E., Hoppenbrouwers, S., & Hengeveld, S. (2012). Abstract reasoning in collaborative modeling, *45th Hawaii International Conference on System Sciences* (pp. 170-179). IEEE.

Wise, A. F., & Schwarz, B. B. (2017). Visions of CSCL: Eight provocations for the future of the field. *International Journal of Computer-Supported Collaborative Learning*, *12*, 423-467.

Xiang, M., Zhang, J., & Li, Y. (2024). Understanding knowledge convergence in a cross-cultural online context: an individual and collective approach. *Proceedings of the 14th Learning Analytics and Knowledge Conference* (pp. 779-784).

Zheng, L., Chen, N. S., Huang, R., & Yang, K. (2014). A novel approach to assess collaborative learning processes and group performance through the knowledge convergence. *journal of computers in education*, *1*, 167-185.

Zheng, L., Zhen, Y., Niu, J., & Zhong, L. (2022). An exploratory study on fade-in versus fade-out scaffolding for novice programmers in online collaborative programming settings. *Journal of Computing in Higher Education*, *34*(2), 489–516.

Zottmann, J. M., Stegmann, K., Strijbos, J. W., Vogel, F., Wecker, C., & Fischer, F. (2013). Computer-supported collaborative learning with digital video cases in teacher education: The impact of teaching experience on knowledge convergence. *Computers in Human Behavior*, *29*(5), 2100-2108.

Acknowledgments

This work has been partly supported by the Grants-in-Aid for Scientific Research (NOs. 21K02752, 21K18527 and 22K02951) by MEXT (Ministry of Education, Culture, Sports, Science and Technology) in Japan

 © ISLS

International Society of the Learning Sciences

Small Arguments, Big Ideas: Fostering Children's Collaborative Reasoning with LLM Agents

Hecong Wang, University of Rochester, hwang99@ur.rochester.edu
Meng Wang, University of Rochester, mwang106@ur.rochester.edu
Erfan Farhadi, University of Rochester, efarhadi@ur.rochester.edu
Asad Shahab, University of Rochester, ashahab4@u.rochester.edu
Yuanzhu Li, University of Rochester, yli284@u.rochester.edu
Haozheng Du, University of Rochester, hdu6@ur.rochester.edu
Carolyn Rosé, Carnegie Mellon University, cp3a@andrew.cmu.edu
Hangfeng He, University of Rochester, hangfeng.he@rochester.edu
Zhen Bai, University of Rochester, zhen.bai@rochester.edu

Abstract: Collaborative reasoning is at the core of critical thinking, problem solving, and effective communication, with growing attention being devoted to developing pedagogical and technological supports that enhance it. While many existing conversational agents (CAs) for education emphasize comprehension or factual recall during reading, the use of advanced reasoning behaviors, such as persuasion and inquiry, may further foster children's engagement in collaborative reasoning. Based on recent advancements in large language models (LLMs), we created CAs that engage children in collaborative reasoning through argumentation during book-reading activities, by operationalizing three dialogue types from Walton's framework of dialogue and argumentation: information-seeking, persuasion, and inquiry. We conducted a four-day exploratory study with 18 8-year-olds at a literacy-focused summer camp. Our findings revealed that while asking questions can support children's collaborative reasoning, incorporating even small amounts of persuasion and inquiry elicits significantly more reasoning behaviors, especially explanations, clarifications, and reflections, from the children.

Introduction

Collaborative reasoning is foundational to learning, enabling individuals to jointly develop their knowledge and understanding (Mercer, 1995) through exchanging information, resolving conflicts, and constructing shared understanding (Walton, 1998; Walton & Krabbe, 1995). It is well studied in learning science for fostering deeper engagement, critical thinking, and meaningful learning experiences (Cui & Teo, 2021; Mercer, 1995).

Dialogic reading (DR) provides a rich environment for collaborative reasoning in early education, through interactive and reflective reading practices between children and adults (Pillinger & Vardy, 2022). It helps children develop critical socio-emotional skills (Fu et al., 2022; Pillinger & Vardy, 2022; Sethi & Jain, 2024; Whitehurst et al., 1994), which are strongly associated with healthy development and success both within and beyond the classroom (Newman & Dusenbury, 2015; Slovák & Fitzpatrick, 2015).

With the advancement of Large Language Models (LLMs), emerging research identifies new and promising opportunities for Conversational Agents (CAs) in engaging children in productive DR activities (Chen et al., 2025; He et al., 2025; Xu et al., 2022; Xu & Warschauer, 2020; Zhang et al., 2022). However, existing reading CAs primarily rely on interactions driven by information-seeking questions (e.g., Zhang et al., 2022 and He et al., 2025) and scripted task-oriented responses (e.g., Hu et al., 2024), leaving a critical gap in leveraging the full spectrum of collaborative reasoning behaviors, such as posing challenges, offering rebuttals, seeking rationales, and providing explanations (Gillies, 2016; Tegos & Demetriadis, 2017).

This study examines the impact of CAs with different dialogue types, specifically information-seeking, persuasion, and inquiry, on engaging children in collaborative reasoning during DR. These dialogue types are informed by the Waltonian framework of dialogue and argumentation (Walton, 1998; Walton & Krabbe, 1995) and are implemented by offering different kinds of responses (i.e., information-seeking, persuasion, and inquiry response types) that each embodies its distinguishing characteristics. This theoretical framework, originating in philosophy and formal argumentation, has been successfully applied to pedagogical agents (Norman et al., 2003; Sklar & Azhar, 2012) and, more recently, to classroom discourse analyses (Foster et al., 2022; Rapanta & Christodoulou, 2022). We conducted a four-day exploratory study involving 18 8-year-old children in a literacy-focused summer camp to interact with four CA prototypes: three that prioritize each response type separately, and a fourth that more freely mixes all three types based on the conversational flow. Our initial analyses, leveraging multiple existing behavior coding frameworks implemented via human-validated automated pipelines and qualitative analyses, show that CA prototypes met design goals and that SOTA LLMs can deliver argumentation-

 © ISLS

based reasoning moves. Persuasion and inquiry responses elicited deeper reasoning behaviors, including question-asking, explanations, and metacognitive reflection. Notably, the mixed condition, despite being predominantly information-seeking with some offering of persuasion (8.2%) and inquiry (9.5%) responses, led to significantly more student justifications and explicit reasoning.

Our main contributions are 1) an initial exploration of the effects of different Waltonian dialogue types on children's collaborative reasoning behavior during DR; 2) CAs implemented using two distinct architectures that offer information-seeking, persuasion, inquiry, and mixed dialogue types; and 3) a demonstration of how LLMs can be effectively used for conversational coding analyses.

Related work

Story-reading conversational agents

DR emphasizes the co-construction of knowledge through meaningful, reciprocal dialogue between readers, fostering critical thinking and active engagement, especially for socio-emotional learning (Wegerif, 2007). Xu and Warschauer (2019) provide an early exploration of the area, demonstrating the promising effects of CAs in engaging children with DR activities through the use of open-ended questions, praise, and explanations. There is emerging work that utilizes LLMs in building CAs for DR. Zhang et al. (2022) introduced StoryBuddy, which enhanced the experience by leveraging a fine-tuned LLM-based CA to dynamically generate questions that were more engaging, logically connected, and aligned with educational objectives. More recently, StoryPal by He et al. (2025) operationalized two DR frameworks (i.e., PEER and CROWD) through prompt engineering, enabling LLMs to generate framework-informed questions, dynamic scaffolding, and contextual feedback. While information-seeking is an important and effective DR practice, it remains unclear how CAs incorporating more advanced collaborative reasoning behaviors would impact children. This work offers an early exploration of the impacts of CAs that prioritize behaviors such as information-seeking, persuasion, and inquiry, as well as offering a natural mixture of all, in supporting collaborative reasoning for children in DR activities.

Collaborative learning and the Waltonian framework

Walton's framework of dialogue and argumentation (Walton, 1998; Walton & Krabbe, 1995) provides a normative theory for understanding collaborative reasoning. It classifies dialogic argumentation into seven dialogue types—persuasion, negotiation, inquiry, deliberation, information-seeking, eristics, and mixed—each defined by its initial conditions, goals, and behavioral rules, which can be further refined into subtypes (Walton & Krabbe, 1995). In educational contexts, information-seeking, persuasion, and inquiry dialogues are particularly relevant, aligning with information exchange, where learners clarify and share knowledge; conflict resolution, where individuals negotiate competing perspectives to reach consensus; and co-construction of meaning, where individuals synthesize ideas to build shared understanding (Foster et al., 2022; Rapanta & Christodoulou, 2022). Walton's framework has practical applications in the design of CAs and classroom interaction analyses, helping categorize teacher-student and peer argumentation in STEM education to foster critical thinking (Foster et al., 2022; Rapanta & Christodoulou, 2022). It guides systems that engage learners in structured argumentation, promoting deeper engagement and reasoning skills (Garrison et al., 1999; Reed, 2021; Sklar & Azhar, 2012). Despite the significant body of work, theory-based argumentation-driven CAs remain underexplored, as highlighted by Castagna et al. (2024). Our work aims to address this gap by creating CAs that operationalize Walton's framework of dialogue and argumentation, chosen for its generalizability compared to DR-specific frameworks, such as PEER and CROWD (Pillinger & Vardy, 2022), and its flexibility for casual conversation over other formal argumentation theories, like Toulmin's method (Toulmin, 2003).

CA design and implementation

We developed four CA prototypes operationalizing Walton's framework of dialogue and argumentation by offering information-seeking, persuasion, inquiry, and a natural mixture of all three response types. The first three CA prototypes were designed as controlled implementations of the three dialogue types that allow us to gain an initial understanding of their strengths and limitations in influencing collaborative reasoning during DR. Using more sophisticated techniques, namely ReAct (Yao et al., 2023), we built the fourth prototype to explore how effectively SOTA LLMs can engage children in collaborative reasoning, under the Waltonian framework.

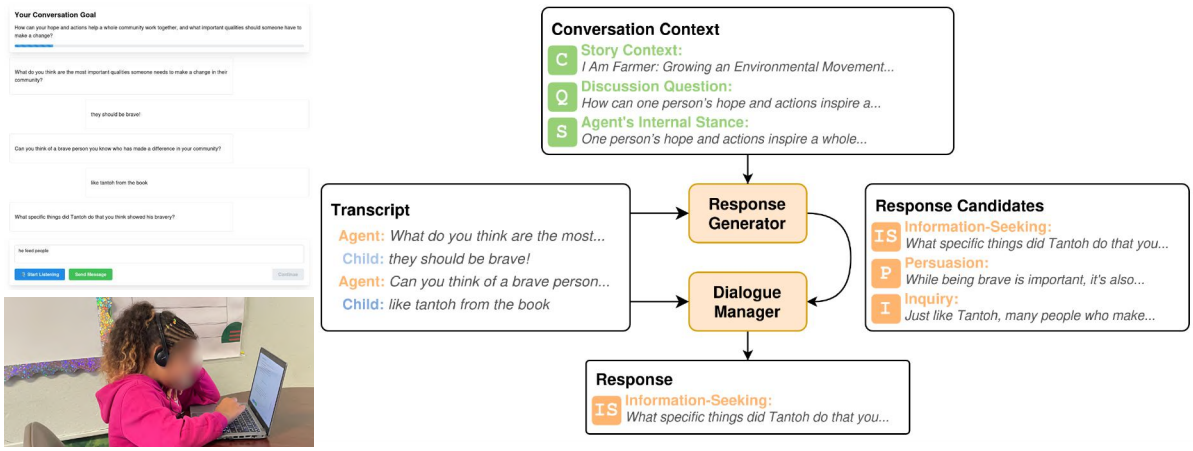
Agent prototyping

The CAs are designed to engage children in DR activities by incorporating the *story context* (i.e., reading materials), an initial *discussion question* (e.g., *How can one person's hope and actions inspire a whole community*

to work together for change?), and an *agent's internal stance* that guides the agent's own reasoning (e.g., *One person's hope and actions inspire a whole community when they combine genuine passion, perseverance, and an inclusive approach that respects and uplifts others*), which could be automatically generated or authored by educators (1). The user interface, shown in Figure 1, resembled a standard chat environment similar to ChatGPT, with the discussion question and a progress tracker displayed at the top to remind students of the question and their progress through the conversation. The two-stage generation approach is illustrated using the example conversation shown. To accommodate participants' diverse literacy levels, the interface includes speech-to-text for child input and text-to-speech for the agent's responses.

Figure 1

System Overview with Interface, Sample Transcript, and Generated Responses



Generation of information-seeking, persuasion, and inquiry responses

To highlight the characteristics of each dialogue type, we adopted a two-stage generation approach common among CA designs, which decomposes the response generation process into a response generator and a dialogue manager (Ramesh et al., 2017). This approach allows for the separation of concerns and precise control over when each response type is offered. We adopted the zero-shot prompting approach for the response generator by providing descriptions of the characteristics of each response type based on Walton's framework (2). An iterative refinement process was employed to adjust the descriptions in the prompt, ensuring the generated responses were more closely aligned with the expected response type. Each CA prototype is then associated with a simple rule-based dialogue manager that prioritizes the target response type. Similar to prior work, such as ContextQ (Dietz Smith et al., 2024), the dialogue managers assess the appropriateness of the generated response, as judged by an LLM. When the response is deemed inappropriate for the conversational context, the manager defaults to LLM-based free response generation for that turn.

Generation of a mixture of the three response types

While the straightforward two-stage generation approach helps deliver and amplify the characteristics of each dialogue type suitable for a controlled experiment, it may undermine the natural flow of human dialogue, thus lacking resemblance to authentic collaborative reasoning. Therefore, we introduced a fourth CA prototype that aims to include a mixture of the three response types while maintaining the natural flow of conversation. We adopted the ReAct prompting technique (Yao et al., 2023), which is centered on the think-then-act process that demonstrates effective dialogue reasoning ability. The ReAct-based agent implements the following reasoning pattern. At the start of each turn, the agent first formulates its current *agenda*, which specifies the immediate conversational goal (e.g., *introduction itself*, *collaborating on a question*, or *completing an unfinished sub-dialogue*). The *obligation* component tracks rules or prior commitments that constrain how the agent should act with respect to the *agenda*. The agent also engages in an internal *thought-critique-revise* process, reasoning about how best to fulfill its agenda, given its *obligations*. Finally, the agent determines the most appropriate *dialogue* move from the theory and generates a *response*. As a form of working memory, we also allow the agent to create *comments*, which are tracked throughout the conversation.

Technical evaluation

We used a manipulation check inspired by Govers et al. (2024) to validate the effectiveness of the proposed zero-shot response generator in generating information-seeking, persuasion, and inquiry response types. We first curated a small human-agent corpus by having two authors interact with the mixed-condition CA prototype, each performing four free-form conversations lasting at least 15 turns. We randomly sampled five agent responses from each of the eight conversations (approximately one-third of all agent responses) and applied the zero-shot generation prompts to produce multiple alternative responses. In total, 240 alternative responses were generated and independently coded by two authors. After reaching good inter-rater reliability by training on three dialogues, two authors independently coded all alternative responses, and an overall Cohen's Kappa of $\kappa \approx 0.80$ was reached, indicating substantial agreement between the coders. After resolving any conflict, we assessed the effectiveness of the generation prompts by measuring the agreement between the human-perceived response type and the prompts' intended response type. We report the agreement using precision, recall, and F1 scores, as shown in Table 1.

Table 1
Technical performance evaluation results

Response Type	Precision	Recall	F1
Information-seeking	100%	93%	96%
Persuasion	53%	95%	68%
Inquiry	90%	65%	76%

Investigating persuasion responses

In light of the lower performance of the persuasion responses, we manually reviewed the generated error cases to understand their common error modes. During our review, we noticed that at certain points in the conversations, a persuasive response is unlikely to be offered because the human never established a debatable statement to be contested. We thus hypothesize that *persuasion responses can be offered selectively only at points in a conversation where debatable claims exist*. To verify our hypothesis, two authors jointly coded each turn of the conversations as debatable or not. In the end, only six out of the total 40 instances were deemed to be conversational turns in which a persuasion response could reasonably be offered. When conditioned on this dimension, we found the F1 score for the persuasion-generation prompt to be 70% in debatable cases and 50% in non-debatable cases, which aligns with our hypothesis (3).

Research method

We conducted an exploratory human-subject study using the CA prototypes to explore two research questions:

RQ1. How to build CAs that can engage children in collaborative reasoning?

RQ2. What are the strengths and limitations of different dialogue types for supporting collaborative reasoning during DR activities?

Study design

The study consists of four conditions, each corresponding to a CA prototype that provides a target dialogue type: information-seeking, persuasion, inquiry, and mixed. Because student attendance varied across days, the study condition was randomly assigned to participants for each day's session. The randomization was designed to ensure that each participant was exposed to each condition at most once, and the conditions were balanced across days, gender groups, and reading materials to the greatest extent possible. In addition to the CA conditions, each day also included a child-child baseline condition in which two participants were randomly selected (without repetition) to discuss the same question with one another.

Each day, a new book was introduced to the participants. The research team created the discussion questions for each book in consultation with the summer camp organizers to ensure relevance to the camp curriculum and age-appropriateness. We used ChatGPT to generate the agent's internal stances for each question based on the teacher's curriculum learning objectives of the book.

Participant

A total of 18 students (10 males, 8 females), all aged 8, participated in the study. Most participants identified as Black or African American and native English speakers. Students exhibited varying levels of prior experience with digital technology, including using laptops, but reported minimal familiarity with LLMs such as ChatGPT.

 © ISLS

Informed assent was obtained prior to the study, and the study was approved by the Institutional Review Board of the academic institute.

Study procedure

The study was conducted over four consecutive days at a literacy-focused summer camp. Prior to the study, one of the authors led an introductory session designed to familiarize participants with the concepts of LLMs and CAs. Participants were invited to have a free-interaction session with the CA prototypes, and feedback from the session was incorporated into the interface.

Each study day, participants took part in a 30-minute study session, following a guided reading session of a new book led by teaching assistants from the summer camp. The study session includes four stages:

1. *Pre-study test* - Participants were asked to answer a pre-defined discussion question about the book.
2. *Collaborative reasoning* - Participants are invited to discuss the same question with one of the CA prototypes. To ensure sufficient interaction, each conversation was required to be at least 15 turns long—a threshold derived from observations during the introductory session.
3. *Post-study test* - Participants were asked to answer the same discussion question after CA interactions.
4. *Usability questionnaire* - Participants completed a short Likert-scale usability questionnaire about their interaction with the CA.

Data collection

We collected pre- and post-study tests using an open-ended questionnaire. Full conversation transcripts and internal system logs were recorded, including response start and end times, dialogue manager decisions, and the model's internal reasoning traces for the ReAct-based CA. We used the Chatbot Usability Questionnaire (CUQ) (Holmes et al., 2019), a 16-item Likert-scale questionnaire developed for evaluating chatbot usability, to assess participants' perceptions of each CA prototype. The child-child baseline condition was video-recorded.

Data analysis

Preprocessing and screening

In total, we collected 53 child-agent dialogues. One author manually inspected and corrected all participants' utterances due to imperfect speech-to-text performance in natural classroom settings. Following the manual screening process, we removed 11 child-agent dialogues that were either fewer than three turns or largely off topic, resulting in 42 valid conversations for analysis. Additionally, 29 completed usability questionnaires and 26 matched pairs of pre- and post-study tests were collected. For the child-child baseline condition, two out of four sessions were retrained for analysis. The session on day one was excluded due to poor audio quality, and the session for day four was excluded because one participant had not completed the assigned reading prior to the discussion. The average number of responses per conversation is 12.4 ± 4.0.

Coding of verbal interaction

For agent utterances, we developed a lightweight classifier for information-seeking, persuasion, and inquiry, trained on 480 manually coded agent responses from the system's technical evaluation phase. We utilized SentenceTransformers (Reimers & Gurevych, 2019), specifically the *all-mpnet-base-v2* model provided by the library, to obtain high-dimensional sentence embeddings. We then trained a logistic regression classifier from Scikit-learn (Pedregosa et al., 2011) to predict the corresponding response type (i.e., information-seeking, persuasion, or inquiry). We evaluated the classifier through five-fold cross-validation, stratified by response type to maintain class distribution, achieving a mean accuracy of 83.1% and a weighted F1 score of 82.9%.

For child utterances, we adopted two established coding schemes of collaborative reasoning behaviors in learning to obtain comprehensive perspectives on the impact of different dialogue types. The first scheme, developed by van Boxtel et al. (2000), contains 11 categories of communicative functions for collaborative reasoning: *arguments*, *evaluations*, *questions*, *requests*, *proposals*, *confirmations*, *rejections*, *repeats*, *orders*, and *off-task* utterances (van Boxtel et al., 2000). The second scheme, developed by Rathod and Murthy (2022), contains nine categories with more detailed perspectives regarding reasoning-oriented moves: *clarify*, *support*, *question posing*, *recognize*, *propose*, *defend*, *reject*, *evidence*, and *causal talk* (Rathod & Murthy, 2022).

To develop automated classifiers for these two coding schemes, we sampled four conversation transcripts from the screened corpus, one from each CA condition. Two authors independently annotated all children's utterances at the sentence level using both coding schemes. A Cohen's Kappa of $\kappa \approx 0.70$ was reached for each

 © ISLS

scheme, indicating substantial agreement. Discrepancies were jointly resolved, yielding a dataset of 168 labeled samples (84 per coding scheme).

We implemented a zero-shot LLM classification pipeline for the two coding schemes, as the previous embedding-based approach yielded poor results. Two prompts were designed following a consistent template: 1) a high-level task description, 2) the coding manual, 3) special notes to clarify ambiguous distinctions, 4) formatting instructions, and 5) the full conversation history ending with the target child's utterance. We iteratively refined the prompts, focusing on the special notes section, to reduce confusion between similar codes. Finally, using GPT-5 (i.e., *gpt-5-2025-08-07*), the two zero-shot prompts achieved an accuracy of 80.5% (weighted F1 of 80.4%) for the van Boxtel et al. (2000) scheme and an accuracy of 77.8% (weighted F1 of 76.9%) for the Rathod and Murthy (2022) scheme.

Results

RQ1. How to build CAs that can engage children in collaborative reasoning?

To evaluate whether the CA prototype in each condition generated the intended dialogue behaviors, we analyzed the distribution of CA response types during child-agent interaction. As shown in Table 2, the CAs in information-seeking, persuasion, and inquiry conditions predominantly produced their intended response types (i.e., over 50%), indicating that the dialogue managers generally succeeded in guiding the CAs' conversational behavior. The mixed-condition CA is predominantly information-seeking, with comparatively higher offering of persuasion and inquiry responses than the information-seeking CA.

Table 2

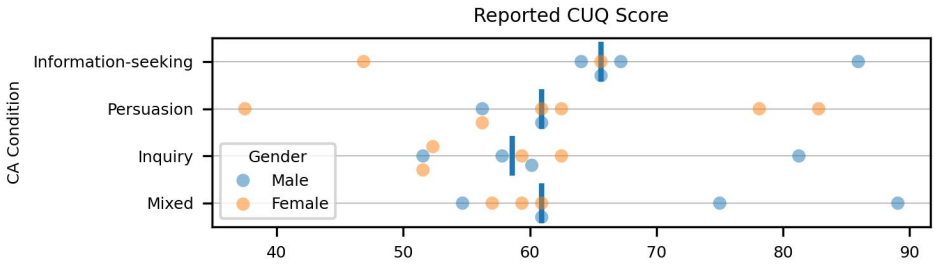
Observed Response Type Distributions Per CA Condition

CA Condition	Information-seeking	Persuasion	Inquiry
Information-seeking	98.8%	1.2%	0.0%
Persuasion	12.5%	57.1%	30.4%
Inquiry	27.1%	4.8%	68.1%
Mixed	82.3%	8.2%	9.5%

Figure 2 presents the distribution of CUQ scores across the four CA conditions, with the median for each condition indicated. Overall, participants rated the agents as moderately to highly usable, with scores ranging between 50 and 75 on the 100-point scale. Among the four conditions, the information-seeking agents received the highest overall median score, followed by the mixed and persuasion conditions, with the inquiry condition showing a slightly lower median score.

Figure 2

Chatbot Usability Questionnaire (CUQ) results

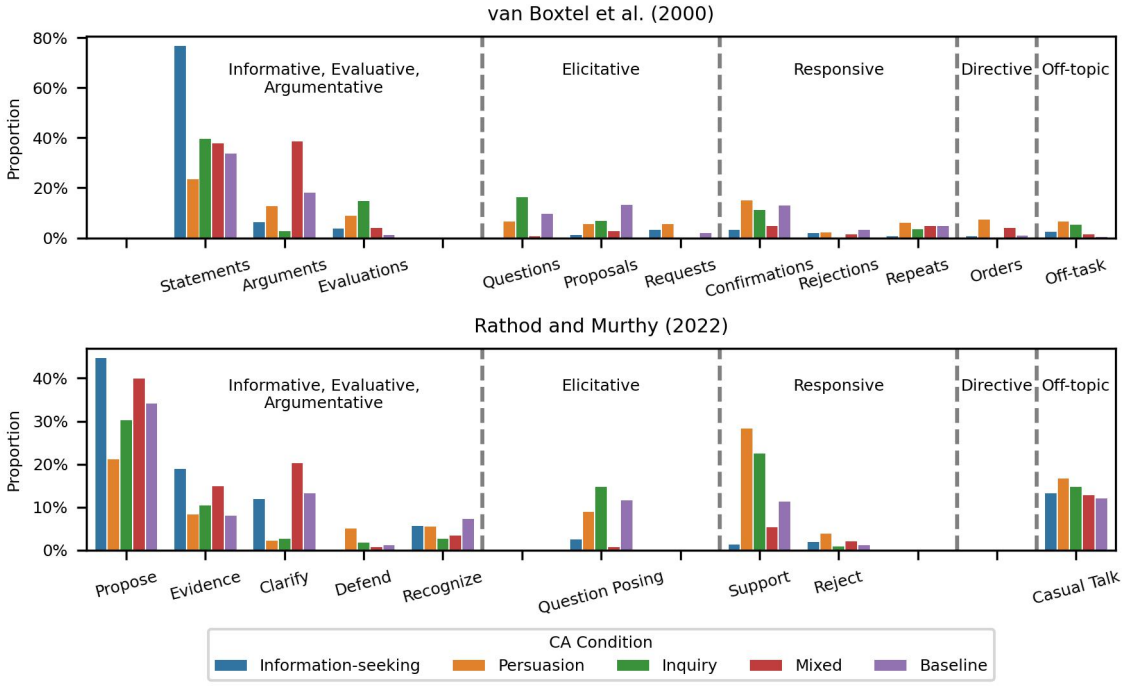


RQ2. What are the strengths and limitations of different dialogue types for supporting collaborative reasoning during DR activities?

Figure 3 shows the distribution of children's dialogue moves across CA conditions and the human baseline, coded using the coding manual proposed by van Boxtel et al. (2000) and Rathod and Murthy (2022). The dialogue moves are grouped by higher-level function and aligned across the two coding manuals. Clear differences emerged in how distinct CA conditions shaped the structure of the student's interactions.

 © ISLS

Figure 3
Collaborative Reasoning Move Distributions Grouped Per CA Condition



Qualitative evaluation results of the pre- and post-test responses are summarized in Table 3, where: 1) *Pre* > *Post* indicates a decline in the response's quality (i.e., reasoning, elaboration, and relevance); 2) *Pre* = *Post* indicates the two responses are identical to or paraphrases of each other; 3) *Pre ≈ Post* indicates two different responses of similar quality (i.e., depth, elaboration, and relevance) were offered; and 4) *Pre* < *Post* indicates improvements in depth, elaboration, or relevance.

Table 3
Pre- and post-test results

CA Condition	Pre < Post	Pre ≈ Post	Pre = Post	Pre > Post
Information-seeking	50.0%	0.0%	33.3%	16.7%
Persuasion	16.7%	33.3%	33.3%	16.7%
Inquiry	28.6%	28.6%	28.6%	14.3%
Mixed	42.9%	14.3%	14.3%	28.6%

Discussion

Information-seeking CA We observed the highest proportion of *Statements* and substantial *Evidence* moves from the children, indicating productive knowledge sharing during the information-seeking CA condition. This observation aligns with existing literature that scaffolding questions support children's generation of claims and recall of evidence (Dietz Smith et al., 2024; He et al., 2025; Xu & Warschauer, 2019; Zhang et al., 2022). However, participants' collaborative reasoning behaviors are confined to these few dialogue moves.

Persuasion and inquiry CAs Compared to the information-seeking and mixed conditions, the persuasion and inquiry CAs were able to elicit a broader range of collaborative reasoning behaviors but were less engaging for children. The inquiry condition saw more *Question*, *Evaluation*, and *Confirmation* moves, indicating exploration and self-monitoring with limited argumentative depth. The persuasion condition got a more balanced mix of *Support*, *Defend*, and *Reject* moves. However, both conditions had lower CUQ scores than the information-seeking condition (see Figure 2). Human inspections of the responses found that agent responses were less responsive to participants (i.e., more responding to children's questions) yet more persistent in pursuing persuasion or inquiry, often producing longer, lecture-like turns that could have disrupted natural conversational flow and reduced engagement.

Mixed CA Participants in the mixed CA condition exhibited the most *Argument* and *Clarify* moves, indicating explicit reasoning and repair to externalize and share their internal thoughts. The condition's CUQ score was similar to that of the persuasion condition, and human inspection revealed that the responses typically followed a three-part structure: 1) acknowledgment of the child's contribution, 2) introduction or extension of ideas by the agent, and 3) a question to further engage the child. This structure is outlined in the agent's prompt.

The dialogue managers generally produced the intended response types, with the mixed agent naturally blending strategies. Our analysis suggests two design principles for improving conversational naturalness and engagement: 1) *Adaptive and Contextual Blending of Dialogue Strategies.* The mixed condition, while predominantly information-seeking, incorporated persuasion (8.2%) and inquiry (9.5%) responses, prompting significantly more *Argument* and *Clarify* moves from children than the persuasion or inquiry agents. Persuasion and inquiry strategies were most effective when applied selectively and in response to conversational context, rather than through fixed or uniform patterns. 2) *Keep Responses Concise and Interactive.* Persuasion and inquiry agents often produced long, lecture-like turns associated with lower usability and fewer reasoning-oriented child moves. Breaking content into micro-turns—short prompts, comprehension checks, and quick follow-ups—helps maintain conversational rhythm and engagement (e,g., DAPIE by Lee et al., 2023).

Limitations and future work

This study has several limitations, and the results should be considered preliminary. 1) the sample was small (i.e., 18 participants) and drawn from a single summer camp; variable attendance led to uneven exposure across conditions, and only one human baseline per day limited peer comparisons; 2) the child-child baseline condition included only male participants due to practical constraints; and 3) the pre- and post-responses were brief because of time limitations in the natural classroom environment. Moreover, the current implementation of the three Waltonian dialogue types represents a simplified interpretation of the theory, realized through turn-level response types rather than episode-level sub-dialogues. Future work will expand the population sample and ensure even exposure across CA conditions using within-subject study designs. We also seek to address the limitations of the current CA prototypes and develop more advanced realizations of the Waltonian dialogue types through adaptive dialogue policies that detect debatable moments, dynamically mix dialogue types, and control response length via micro-planning; gathering and incorporating stakeholder feedback and pedagogical best practices. Finally, we also plan to leverage additional measures (e.g., delayed post-tests, transfer tasks, discourse-based reasoning rubrics, and engagement measures) to deepen our understanding of the underlying mechanisms linking collaborative reasoning with learning processes and outcomes.

Conclusion

This work examined how four CAs (i.e., information-seeking, persuasion, inquiry, and mixed) can support children's collaborative reasoning during DR activities. The information-seeking agent was rated highest for usability and engagement, while the persuasion and inquiry agents elicited broader reasoning behaviors but often produced longer, less interactive turns. The mixed agent, blending multiple strategies, prompted the most reasoning-oriented moves from children, suggesting the benefits of adaptive and context-sensitive strategy use. Overall, the findings indicate that conversational agents should balance structure with flexibility and maintain concise, interactive exchanges to better foster children's reasoning and engagement. However, given the small sample and simplified dialogue models, these results should be viewed as preliminary but point toward promising directions for designing more natural and pedagogically effective child-agent interactions.

Endnotes

(1) We found that providing the agent with an internal stance helps ensure the agent's own logical consistency throughout conversations, and providing a stance that covers various perspectives of the discussion questions enables the agent to have in-depth, multi-perspective discussions during conversations.
(2) We initially explored the approach of generating each response type via a dedicated prompt. We later discovered that simultaneously generating all three responses yields a significant performance improvement (up to 13% relative improvement in F1 score).
(3) This finding motivated us to include the free-response fallback in the dialogue managers used in the human-subject study.
(4) We provide a spreadsheet with our coding schemes and example dialogue snippets here: https://tinyurl.com/y6dtxzdk.

 © ISLS

References

Castagna, F., Kökciyan, N., Sassoon, I., Parsons, S., & Sklar, E. (2024). Computational argumentation-based chatbots: a survey. *Journal of Artificial Intelligence Research*, *80*, 1271-1310.

Chen, J., Tang, M., Lu, Y., Yao, B., Fan, E., Ma, X., ... & He, L. (2025, April). Characterizing llm-empowered personalized story reading and interaction for children: Insights from multi-stakeholder perspectives. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (pp. 1-24).

Cui, R., & Teo, P. (2021). Dialogic education for classroom teaching: A critical review. *Language and education*, *35*(3), 187-203.

Dietz Smith, G., Prasad, S., Davidson, M. J., Findlater, L., & Shapiro, R. B. (2024, June). ContextQ: Generated questions to support meaningful parent-child dialogue while co-reading. In *Proceedings of the 23rd annual ACM interaction design and children conference* (pp. 408-423).

Foster, J., Gillespie Schneider, J., Franco, L., Zhuang, Y., Crawford, B., & Conner, A. (2022). Categorizing Classroom-based Argumentation in Elementary STEM Lessons: Applying Walton's Types of Argument Dialogue. *Journal of Research in STEM Education*, *8*(2).

Fu, Y., Michelson, R., Lin, Y., Nguyen, L. K., Tayebi, T. J., & Hiniker, A. (2022). Social Emotional Learning with Conversational Agents: Reviewing Current Designs and Probing Parents' Ideas for Future Ones. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *6*(2), 1-23.

Garrison, D. R., Anderson, T., & Archer, W. (1999). Critical inquiry in a text-based environment: Computer conferencing in higher education. *The Internet and Higher Education*, *2*(2-3), 87-105.

Gillies, R. M. (2016). Dialogic interactions in the cooperative classroom. *International Journal of Educational Research*, *76*, 178-189.

Govers, J., Velloso, E., Kostakos, V., & Goncalves, J. (2024, May). AI-driven mediation strategies for audience depolarisation in online debates. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (pp. 1-18).

He, K., Gastón-Panthaki, A., Zhuo, D., Munsey, J., Zhang, M., & Warschauer, M. (2025). StoryPal: Supporting Young Children's Dialogic Reading with Large Language Models. In *Proceedings of the 24th Interaction Design and Children* (pp. 494-511).

Holmes, S., Moorhead, A., Bond, R., Zheng, H., Coates, V., & McTear, M. (2019, September). Usability testing of a healthcare chatbot: Can we use conventional methods to assess conversational user interfaces?. In *Proceedings of the 31st European Conference on Cognitive Ergonomics* (pp. 207-214).

Hu, Z., Hou, H., & Ni, S. (2024, June). Grow with your AI buddy: designing an LLMs-based conversational agent for the measurement and cultivation of children's mental resilience. In *Proceedings of the 23rd Annual ACM Interaction Design and Children Conference* (pp. 811-817).

Lee, Y., Kim, T. S., Kim, S., Yun, Y., & Kim, J. (2023, April). Dapie: Interactive step-by-step explanatory dialogues to answer children's why and how questions. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (pp. 1-22).

Mercer, N. (1995). *The guided construction of knowledge: Talk amongst teachers and learners*. Multilingual matters.

Newman, J., & Dusenbury, L. (2015). Social and emotional learning (SEL): A framework for academic, social, and emotional success. In *Prevention science in school settings: Complex relationships and processes* (pp. 287-306). New York, NY: Springer New York.

Norman, T. J., Carbogim, D. V., Krabbe, E. C., & Walton, D. (2003). Argument and multi-agent systems. In *Argumentation machines: New frontiers in argument and computation* (pp. 15-54). Dordrecht: Springer Netherlands.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*(85), 2825–2830.

Pillinger, C., & Vardy, E. J. (2022). The story so far: A systematic review of the dialogic reading literature. *Journal of Research in Reading*, *45*(4), 533-548.

Ramesh, K., Ravishankaran, S., Joshi, A., & Chandrasekaran, K. (2017, May). A survey of design techniques for conversational agents. In *International Conference on Information, Communication, and Computing Technology* (pp. 336-350). Singapore: Springer Singapore.

Rapanta, C., & Christodoulou, A. (2022). Walton's types of argumentation dialogues as classroom discourse sequences. *Learning, Culture and Social Interaction*, *36*, 100352.

Rathod, B. B., & Murthy, S. (2022). What's Worth Solving? An Expert Study to Identify Problem-Finding Strategies within a Socio-Scientific Issue. In *Proceedings of the 15th International Conference on*

Computer-Supported Collaborative Learning-CSCL 2022, pp. 195-202. International Society of the Learning Sciences.
Reed, C. (2021). The Waltonian Foundations of Argument Technology. *IfCoLoG Journal of Logics and Their Applications*, *8*(1), 245–262.
Reimers, N., & Gurevych, I. (2019, November). Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* (pp. 3982-3992).
Sethi, S. S., & Jain, K. (2024). AI technologies for social emotional learning: recent research and future directions. *Journal of Research in Innovative Teaching & Learning*, *17*(2), 213-225.
Sklar, E., & Azhar, M. Q. (2012, September). Toward the Application of Argumentation to Interactive Learning Systems. In *Argumentation in Multi-Agent Systems: 8th International Workshop, ArgMAS 2011, Taipei, Taiwan, May 2011, Revised Selected Papers* (p. 213). Springer.
Slovák, P., & Fitzpatrick, G. (2015). Teaching and developing social and emotional skills with technology. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *22*(4), 1-34.
Tegos, S., & Demetriadis, S. (2017). Conversational Agents Improve Peer Learning through Building on Prior Knowledge. *Educational Technology & Society*, *20*(1).
Toulmin, S. E. (2003). *The Uses of Argument* (2nd ed.). Cambridge University Press.
Van Boxtel, C., Van der Linden, J., & Kanselaar, G. (2000). Collaborative learning tasks and the elaboration of conceptual knowledge. *Learning and instruction*, *10*(4), 311-330.
Walton, D. N. (1998). *The new dialectic: Conversational contexts of argument*. University of Toronto Press.
Walton, D., & Krabbe, E. C. (1995). *Commitment in dialogue: Basic concepts of interpersonal reasoning*. State University of New York Press.
Wegerif, R. (2007). *Dialogic education and technology: Expanding the space of learning*. Boston, MA: Springer US.
Whitehurst, G. J., Arnold, D. S., Epstein, J. N., Angell, A. L., Smith, M., & Fischel, J. E. (1994). A picture book reading intervention in day care and home for children from low-income families. *Developmental psychology*, *30*(5), 679.
Xu, Y., Aubele, J., Vigil, V., Bustamante, A. S., Kim, Y. S., & Warschauer, M. (2022). Dialogue with a conversational agent promotes children's story comprehension via enhancing engagement. *Child Development*, *93*(2), e149-e167.
Xu, Y., & Warschauer, M. (2019, May). Young children's reading and learning with conversational agents. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1-8).
Xu, Y., & Warschauer, M. (2020, June). Exploring young children's engagement in joint reading with a conversational agent. In *Proceedings of the Interaction Design and Children Conference* (pp. 216-228).
Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K., & Cao, Y. (2023, January). ReAct: Synergizing Reasoning and Acting in Language Models. In *International Conference on Learning Representations (ICLR)*.
Zhang, Z., Xu, Y., Wang, Y., Yao, B., Ritchie, D., Wu, T., ... & Li, T. J. J. (2022, April). StoryBuddy: A human-ai collaborative chatbot for parent-child interactive storytelling with flexible parental involvement. In *Proceedings of the 2022 CHI conference on human factors in computing systems* (pp. 1-21).

Acknowledgments

We would like to thank Jeremy Smith, Kelvin Sanchez-Orona, and the Servant leaders and Junior Servant leaders at the Freedom Scholars Learning Center for their generous support, as well as scholars who participated in our study and their parents.

International Society of the Learning Sciences

A Design Space for Dialogic Assessment in CSCL

Rolf Steier, Oslo Metropolitan University, rolf.steier@oslomet.no
Magnus Hontvedt, University of South-Eastern Norway, manh@usn.no
Jacob Davidsen, Aalborg University, jdavidsen@ikp.aau.dk
Charlotte Beal, University of South-Eastern Norway, Charlotte.Beal@usn.no
Morten Oddvik, University of South-Eastern Norway, Morten.Oddvik@usn.no
Marit Solem, University of South-Eastern Norway, Marit.S.Solem@usn.no

Abstract: Dialogic assessment remains remarkably unexplored as a research and design setting in CSCL. We address this gap by developing a design space to support future partnership-based work in our field. Our design space reflects a 1.5-year investigation of dialogic assessment practices and perspectives conducted through literature review work, focus group interviews with practitioners, formal and informal discussions with colleagues and prototype development work with technology designers. First, we frame the design space concept before arguing for the relevance of dialogic assessment as a potential CSCL activity. Then, we present our design space through a cumulative distillation of four relevant perspectives and seven design dimensions. Finally, we conclude by laying out an agenda reflecting how this design space can inform future work and prompt innovative assessment practices.

Introduction

With few exceptions, dialogic forms of summative assessment like oral exams have been perceived in opposition to movements towards digitalization and pedagogical innovation in education. Despite decades of pedagogical transformation efforts towards teaching and learning that is student centered, based on dialogue and collaboration, and supported by technology, the corresponding summative assessment practices have been surprisingly static and constrained (Boud & Bearman, 2022). We have a strong tradition in the learning sciences towards tools, activities, and concepts that support situated learning conversations, but the oral exam as a context for such learning remains almost entirely unexplored. The view of an oral exam as an un-mediated, technology-free, (and often decontextualized) discussion between learner and examiner with little relevance for learning persists. We see evidence of this tension in recent years with the responses to the rapid growth of Large Language Models (LLMs) driven by generative artificial intelligence (GenAI). Chatbots like ChatGPT are disrupting traditional exam methods at all levels of the educational system, enabling students to perform conventional assessments, such as scientific essays, with just a few keystrokes (McGrath et al., 2024). In response, we see renewed interest in "old-fashioned" assessments such as pen and paper tests, and naturally oral exams (Hartmann, 2025). A handful of studies adapt oral exams to online (Akimov & Malin, 2020) and video-based settings (Bayley et al., 2024). However, the main motivation to reconsider oral exams seems to be directly because of a lack of computer-support rather than because of an interest in bringing assessment methods into alignment with current perspectives on dialogic and collaborative learning.

This misalignment underscores a need to develop assessment formats that are fair, systematic, and following teaching approaches that value dialogue, participation, and digital possibilities. *The central premise of our work here is that dialogic assessment should be (re)conceptualized as a computer-supported collaborative learning context* and as a fertile ground for pedagogical and technological innovation.

Dialogic assessment formats facilitate interaction between examiners and students and enable assessing students' active use of knowledge. Unlike traditional assessments, which often test static memorization and reproduction of information (Collins & Halverson, 2018; Pea et al., 2025), dialogic formats assess skills like problem-solving, conceptual application, and adaptive thinking. However, a persistent challenge in dialogic assessment is that dialogue is inherently ephemeral, often leaving students unclear about how their grades are determined. Additionally, examiners, as active participants in the assessment conversation, require support to be consistent in balancing their roles as both participating and evaluating the conversation. This also relates to validity in assessment, which concerns whether the tasks provide trustworthy information about what students know and can do, and whether subsequent interpretations are well justified by theory and empirical data While CSCL research has long explored technology's role in supporting collaborative learning, its potential for dialogic assessment remains underexplored. To date, we have not identified existing tools or practices designed specifically to support dialogic assessment.

It is important to discuss how dialogic assessment, as a summative assessment practice, might be positioned in relation to the notions of collaboration and computer-support central to the field of CSCL. We argue

 © ISLS

that bringing this kind of assessment conversation into alignment with sociocultural learning perspectives invites considering the conversation as a learning activity itself and not merely an evaluation post-learning. The goal of this conversation should be for the student and examiner(s) to collaboratively elicit the student's understanding. This is very much a joint activity with a shared goal of making learning visible. A CSCL activity is built on a triadic structure with a computational artifact mediating the interaction between two or more participants (Ludvigsen & Steier, 2019). We see many opportunities for such computational artifacts to enhance dialogic assessment which build on rich research traditions in CSCL. For example, teacher dashboards have emerged as tools to support teachers in visualizing, monitoring, and adjusting to the complexities of classroom group-work (Matuk, et al., 2019). How might the lessons from these kinds of orchestration tools be adapted to dialogic assessment situations to support examiners in managing the assessment activity in real time? Similarly, scripts, as support structures, have emerged as an important contribution within CSCL to support collaboration (e.g. Noroozi, et. Al., 2013). Could assessment scripts be developed to guide student and examiner in the kinds of knowledge production relevant to an oral exam setting which preserve the possibility for summative evaluation? We can also imagine that findings from CSCL research on learning with simulations (e.g. Hontvedt & Arnseth, 2013) could help create assessment situations that are more authentic and contextualized in real situations. In developing the current design space, we are particularly interested in the ways that perspectives from video- and interaction analysis can inform how examiners systematically interpret the assessment conversation. The goal with the current design space is not to specify all the relevant avenues of exploration within CSCL, but to establish a systematic and structured foundation inviting future work.

When it comes to dialogic assessment, one size does not fit all. We consider dialogic assessment as a *format*, that needs to be actively translated into a concrete assessment *method*. Against this backdrop, this study proposes a *design space* (MacLean et al., 1991; Coelho et al., 2025) for (computer-supported) dialogic assessment. We argue that this design space is necessary for aligning assessment practices with pedagogical goals, practitioner needs, and technological advancements. The overarching agenda is to create a foundation for exploring dialogic assessment like oral exams as a computer-supported activity that is also reflective of theoretical views of learning based in dialogic and collaborative epistemologies. We outline how this space can facilitate collaboration among stakeholders, researchers, and designers. Within this design space, actionable methods tailored to specific subjects and local contexts can be co-created, tested, and eventually implemented with students. We pose the following research question: *How can a design space for dialogic assessment be configured to effectively translate objectives and situational contingencies into actionable computer-supported and collaborative assessment methods?*

Approaches toward a design space for dialogic assessment in CSCL

A design space refers to a conceptual and structured environment created through the configuration of resources, frameworks, and tools that enable the exploration of possibilities, constraints, and dimensions relevant to the design process (MacLean et al., 1991; Coelho et al., 2025). A well-developed design space enables researchers to demonstrate how they navigate challenges, engage with target communities, and deal with potential risks (Lee et al., 2025). A well-functioning design space relies on a set of shared concepts and resources that facilitate productive discussion and framing of goals, challenges, and opportunities for innovation. Through structured dialogue and exploration within this space, participants can communicate effectively and co-construct shared objects (Lee et al., 2025). A design space analysis articulates the basis for design choices and evolves alongside the artifact being developed and can help learning scientists articulate: "Here's what I plan to do today, and here's why I'm making that choice" (Coelho et al., 2025, p.1680). As such, a design space is providing a foundation for reasoned decision-making and iterative refinement.

The dialogic assessment design space builds on a sociocultural perspective on learning and assessment, emphasizing how contexts are collaboratively shaped and re-shaped through the use of conceptual, cognitive, and physical tools (Daniels, 2010). We thus build on the assumption that the specific configuration of dialogic assessment needs to be attuned to each specific course's assessment criteria and objectives. For example, the technological needs as well as ways of eliciting questions would be different in a language class in secondary school, an upper secondary history course, or a pedagogy course in a teacher education program. The design space acts as a dynamic environment that encourages the investigation of unaddressed issues and overlooked perspectives. It is in this iterative process of exploration and reframing that opportunities for innovation emerge. Importantly, treating the creation of the design space as both a part of—and an object of—the research process in this paper, makes the foundation of the design transparent and open for inspection. This helps position the final design solution as not merely a standalone artifact but as a sociocultural and sociotechnical outcome shaped by collaborative practices, contextual influences and technological affordances. In this paper, we develop a design space that comprises four mutually informed perspectives: "institutional," "practitioner," "interaction researcher," and "tool designer" that contribute to enabling a design space for dialogic assessment. This approach underscores

 © ISLS

the importance of design spaces as environments for fostering creativity, critical engagement, and thoughtful problem-solving.

To create a transparent and research-based design space, we draw on several methods: literature reviews of existing oral exam practices and uses of interaction analysis methods, conversations with stakeholders, and focus group interviews with practitioners. Additionally, we draw on insights from the prototyping process of a theory-grounded and practitioner informed digital tool that combines video annotation with AI-based coding and feedback to analyze oral assessment dialogues. This on-going iterative design work involves working with students, educators, researchers, and stakeholders. We briefly specify methods for these various knowledge sources in relation to the four perspectives below. The analysis shows how technology affordances are enacted based on the theoretical and practical knowledge drawn from our investigations. This work represents 1.5 years of systematic investigation into dialogic assessment, and the purpose in presenting this design work is not to put forth a particular design solution but to present the lessons that are learned from these kinds of technology-focused conversations and how they can inform the emerging design space.

The design space for dialogic assessment

Dialogic assessment brings attention to the types of knowledge that emerge through interaction, problem-solving, and joint meaning-making in real time—forms of knowledge often missed in traditional assessment formats like written texts. This aligns with recent learning sciences research, that emphasize the importance of assessment formats that capture interaction and the use of artefacts (Pea et al., 2025). In this section, we outline the design space for dialogic assessment by introducing four perspectives that each contribute valuable and productive insights. By "perspective," we mean a particular lens for interpreting and addressing key considerations in designing this space. For each perspective, we provide an overview, discuss its research foundation, and illustrate its relevance using examples from our work in the Norwegian context. We then identify seven dimensions that guide and structure the exploration of this design space. We use the term dimension as an extension of the spatial metaphor; dimensions describe the design space and allow us to position new assessment methods in relation to one another and in relation to critical questions (Lee et al., 2024).

Key perspectives on dialogic assessment

We describe four perspectives that are understood as cumulative and building on each other. We argue that these perspectives should be involved in design work and prototyping, with the goal of materializing concrete actionable methods suited for situated contexts.

Perspective 1: The institutional assessment practices

Assessment, particularly *summative* assessment methods, are key for students to progress in an educational system. While it is widely accepted that learning involves social and complex processes that develop over time with many possible trajectories, formal assessment situations tend to be individual and involve little interaction with others. Dialogic assessment formats exist but are often less easily standardized, documented, and compared across students and contexts than written tests, which poses challenges for their integration into established assessment systems. Still, assessment is given great authority in educational systems and shapes the students' future opportunities. The summative assessments are usually regulated through policy documents at the school-, national-, and international levels. Therefore, any design process for dialogic assessment must build upon these existing institutional practices and regulatory frameworks to ensure relevance and feasibility. Accordingly, the institutional perspective becomes a vital foundation for designing innovative assessment methods.

To illustrate the importance of this perspective and how it may manifest, we provide an example from the Norwegian context, a particularly relevant case due to its long tradition of dialogic assessment through oral exams. Here, students undergo high-stakes oral and written exams in at least one school subject after year 10 in secondary school and potentially several oral exams in upper secondary school as well as in higher education. The institutional goal of these exams is to assess students' competence within a school discipline, such as Norwegian, mathematics, or science. 'Oral exams' are standard within the education system from secondary school through higher education, especially in subjects like languages and social sciences. However, oral exams are not standardized tests but take the method of a subject conversation between the student and an internal examiner (typically the student's teacher). An external examiner, normally a teacher from another school district, also attends the exam, and in the event of disagreement about the student's grade, the external examiner is given primary rights to decide on the final grade. While there are national guidelines on how to conduct an oral exam, most relate to the administration of the exams, with limited guidance on how teachers should assess students' performance. The examiners have primary responsibility for designing the questions based on curriculum standards, as well as evaluating students' performance, and guidelines concerning the actual examination are

restricted to the rather vague formulation that the examination should provide students with “the opportunities to show as broad competence as possible” (Norwegian Directorate for Education and Training, 2026). Research on oral exams has shown that the examiners’ management of the conversation creates the basis for the student to demonstrate competence (Sikveland et al., 2021; Skovholt et al., 2021; Vonen et al., 2022).

Against this backdrop and the case of Norway, it is evident that the institutional practices and frameworks governing assessments needs to be included in a design space for dialogic assessment. Institutional practices set constraints and opportunities for designing new assessment methods. A design process of dialogic assessments needs to build on and acknowledge the need to include the institutional perspective to see opportunities for variation and innovation. For example, in national guidelines for Norwegian oral exams it is explicitly stated that students should be able to present their work, engage in discussions, and demonstrate an independent stance. However, the use of technologies is not specified in detail within the guidelines, nor is it restricted. Students commonly use tools like PowerPoint, although this practice is not explicitly mentioned. We propose that technologies could be meaningfully integrated into dialogic assessments to enhance the conversation and support both students and examiners. Including institutional perspectives in the design space for dialogic assessment is crucial to identify paths for redesign and refinement in a design space.

Perspective 2: The practitioners’ experience with conducting exams

Building on the consideration of institutional practices and formal guidelines, we now turn to a less documented yet equally important aspect of dialogic assessment: the practitioners’ perspective. Dialogic assessment is inherently relational, relying on the conversation between examiners and students. While general aspects of the assessment process are closely regulated, the actual practices depend heavily on the professional judgment and expertise of examiners. As such, including the practitioners’ perspective and experiences are crucial for creating a design space for dialogic assessment.

In our work in Norway, we have conducted a first step of mapping practitioners’ perspectives through conversations with county representatives, with school leaders, and conducted group interviews with examiners in secondary school and in teacher education. The focus group interviews were conducted with lower secondary school teachers (N=4, 1 hour and 22 minutes over Zoom), secondary school teachers (N=3, 60 minutes) and teacher educators (N=4, 60 minutes). The teachers had experience as examiners in the subjects of Norwegian, Social Studies, English as a Second Language, Natural Sciences, Mathematics and Religion. All teachers were recruited based on previous work affiliations, and the interviews were audio recorded. We then transcribed the interview data and mapped general themes that occurred across participants’ perspectives. At the county level, they acknowledged that the current institutional practices for oral exams need innovation to secure student rights and enable them to formally appeal and request re-assessment. The county we collaborate with will therefore experiment with recording the students’ oral exams. School leaders acknowledge that current practices need innovation but also point to the importance of involving examiners in this process, not new rules imposed top-down. In the focus group interviews, the practitioners in secondary school and teacher education highlighted oral exam situations as conducive for displaying students’ communicative competence in the subject. They acknowledged that oral exams allow assessing students’ analytic ability and the ways in which they engage with knowledge resources, e.g., concepts, images or questions asked. They described that the sensitive nature of oral assessment situations could influence both their mental presence as examiners, and the assessment process. The conversation between examiners after the exams were brought forth as a productive way to ensure a high degree of objectivity for assessing the students’ competence, yet that the oral exam format had limitations: it provides only a "snapshot" of a student's competence. The teachers expressed a need for support that could provide oversight and evaluate students more objectively. Here, they suggested that ‘an extra set of eyes’ could be a meaningful way support the assessment process, and their own training as examiners. Video was suggested as a useful resource in relation to the latter; ‘look at recordings of a few exams afterwords to learn better myself’ together with a ‘teacher council’ for professional development. The teachers tend to focus on the most overarching skills, such as analytical abilities and the capacity to apply theory. Introducing student voices in a more systematic way in the partnership process will be essential. We have had initial conversations with students and are now preparing a survey in two counties to further explore students’ views on fairness, as well as the opportunities and challenges associated with oral exams.

Clearly, the examiners’ perspectives in the design space are key to identifying needs and gaining understanding for the complex, relational and situated practices of doing oral assessment. Engaging with local authorities, school leaders, and educators gives a deeper insight of the practical realities and relational aspects of dialogic assessments. This is in alignment with a partnership-based approach that facilitates innovation, while ensuring that new tools and practices are grounded in the lived experiences of those directly involved in the assessment practices.

Perspective 3: The interaction analyst' strategies for conceiving dialogues

The first two parts of the design space involve institutional practices and participant perspectives that both are key in an initiative to develop a design space to innovate dialogic assessment. We move forward by proposing a specific line of research that has well-developed methods for empirically analyzing dialogues. Interaction analysis (Jordan & Henderson, 1995), and more broadly conversation analysis (Sidnell & Stivers, 2012) and video analysis (Derry et al., 2009), enables the analysis of dialogue as it develops sequentially within social practices. Interaction analysis focuses on the fine-grained details of social interactions as they unfold in time and space, and uses video as the key resource for conducting such analysis. Based on a literature review we conducted on interactive use of video in higher education, findings show that video recordings of talk can become meaningful tools for formative and summative assessment. In short, through a framework synthesis approach (Dixon-Woods, 2011) we mapped qualitative research addressing practices and applications in which video technology was employed as a resource for group learning and assessment. Searches were conducted in Education Research Information Centre (Eric), Web of Science and Academic Premier in June 2023 and January 2025 by using the following parameters: 1. Video (in title), 2. OR "interactive video", 3 AND collaborat* OR interact* OR teaching OR lecturing OR assessment OR feedback OR classroom OR dialogue (in abstract), 4. AND "higher education". These searches resulted in 2578 studies. After removing duplicates and screening abstracts, 92 studies were included in the review. Studies were included if they examined activities where two or more students interactively engaged with video for learning or assessment purposes. Most studies were excluded when solely based on quantitative data, video was only part of research method, focus was on individual learners interacting with video, or interactive use of video not being main phenomenon under study. Various kinds of video-based and interactive learning activities were shown to afford students opportunities to practice, assess and improve their communication and language skills (Chi et al., 2017; Greenblatt & McDonald, 2024; Roza & Rustam, 2023). Students have also reported that time spent watching video recordings of their own presentations helped improve their presentation skills (Ritchie, 2015).

In related work from the Norwegian context, recent conversation analytic studies have shown that the examiner's conduct influences student performance (Sikveland et al., 2021; Skovholt et al., 2021; Vonen et al., 2022). It is shown how oral exams can allow students to collaboratively construct and re-contextualize prior experiences, expanding on them in relation to the broader knowledge structures of a discipline (Hontvedt, & Solem, 2024;). Analysis of the grading conversations between examiners after oral exams shows that assessment tends to be based on vague interpretations of the assessment criteria, with little explicit connection to the students' talk (Solem et al., 2023). This highlights the need to experiment with oral assessment methods that ensure transparency and anchor grading discussions in student performance.

By incorporating interaction analysis perspectives into the design space for dialogic assessment, we can develop methods that are better equipped to capture the complexity of interactions in assessment contexts. This involves recognizing the examiner's role in creating a context that allows students to demonstrate their knowledge. Another key aspect is the "next-turn proof procedure" for investigating how a participant's contribution builds on and responds to the preceding utterance. Lastly, but perhaps most importantly, interaction analysis has demonstrated the value of creating detailed transcripts and visualizations of conversations to enable thorough analysis (Davidsen & Steier, 2024; Hall & Stevens, 2015; Shapiro et al., 2025) and marks a possible avenue for innovating dialogic assessment.

Perspective 4: The CSCL designer's eye for technology affordances

The fourth perspective builds on the previous three and emphasizes the potential contribution of technology to support dialogic assessment, through the lens of a CSCL designer. We use this term to include knowledge on the affordances of technology to enhance collaboration and interaction in learning contexts. As a research domain, CSCL is particularly well-positioned to contribute to methods and tools that emphasize both "discourse and interaction" and the "interactional mediation" of technology in teaching and learning (Stahl & Hakkarainen, 2021).

In our case in Norway, we have collaborated with CSCL designers to develop a working prototype: 'the traced dialogic assessment tool'. This work has helped us begin to concretize designer perspectives in a particular case. The tool is implemented as an Electron-based desktop application that combines speech-to-text transcription (using Whisper), natural language processing (through GPT models), and interactive interfaces for transcript editing and annotation. This architecture makes it possible to transform ephemeral oral exchanges into durable representations that can be inspected collaboratively by examiners. Rather than treating oral exams solely as individual performances to be scored, the tool foregrounds them as jointly constructed interactions. Within the application, examiners can import recordings, generate and import transcripts, and review them in synchrony with the original audio. A waveform visualization supports navigation through the recording, while transcripts remain coupled with the audio to preserve features such as prosody and hesitation that are often significant in dialogic

assessment. Using a combination of GPT-based analysis and locally implemented fuzzy matching, the tool highlights transcript segments in which students appear to demonstrate understanding of key concepts defined by the examiner or identified in the syllabus. These suggestions can then be accepted, rejected, or modified, ensuring that responsibility for evaluation remains with the human examiner. At the same time, an integrated editor allows examiners to insert their own annotations, highlight relevant passages, and build layered records that combine AI-generated suggestions with human judgment. A central affordance of the tool is its configurability. Examiners are able to define the key ideas and assessment criteria that guide the analysis, set thresholds for the confidence of AI-generated suggestions, and adapt prompts to specific disciplinary contexts. In this way, the prototype accommodates local differences in curriculum design and assessment culture. Beyond the moment of examination, the tool also enables export of annotated transcripts and assessment reports, offering possibilities for feedback, documentation, and reflection. These outputs can make the rationale for grading more transparent to students and create new opportunities for formative use of summative assessments.

In our pilot work, a class of 20 information science students conducted mock oral exams using the traced dialogic assessment tool and then reflected on their experiences in group presentations. Through these presentations, participants suggested new configurations, visualizations, and annotation functions, illustrating concrete directions for further development of the tool. Another take-away from these sessions involved students' reflections on their participations in the exam – with the computational resources they started to identify patterns and traces in their dialogic participation as individuals and as a group. This process enabled us to collaboratively envision possibilities for traced assessment and to identify priorities for future work.

For this study, the purpose of the prototyping and design work is to provide insight into (CSCL design) perspectives and not necessarily to advance a specific technical solution. Based on this work, we hypothesize that CSCL technologies can be harnessed to make the interpretive process of assessment more systematic, transparent, and accountable, without replacing professional judgment in assessment. This is a timely topic, since recent developments of AI-supported transcription tools, visualization tools, and annotation tools can support analysis of dialogues (Davidsen & Steier, 2024; Shapiro et al., 2025). This includes alternative visualizations of transcripts (e.g. bubble diagrams, turn-taking charts, spatio-temporal diagrams, etc.), learning analytics dashboards or multimodal learning analytics. Furthermore, producing a material artefact from the dialogue can support transparency and object-oriented collaboration that goes beyond the concrete assessment situation. (Damsa, 2014). We carried these insights through to the development of design space dimensions below.

Dimensions of design space for dialogic assessment

By looking across findings from the above perspectives, we inductively and iteratively developed a set of dimensions that can be used to characterize a dialogic assessment method (Table 1). We clarified dimension descriptions and possibilities in relation to existing CSCL frameworks, reducing these to the seven below. As noted above, dialogic assessment in general is seen as a format and these dimensions support the translation work of moving to concrete computer-supported dialogic assessment *methods*. Note that some of the dimensions are linear whereas others are more categorical. For a given method, some dimensions may be more relevant than others. As the field of CSCL has paid so little attention to dialogic assessment contexts, it can be challenging to anchor new design ideas in relation to other emerging designs. These dimensions, as part of the proposed design space, open up useful questions for designing new assessment methods and their associated tools. We can also identify fruitful areas of exploration by combining dimension possibilities in new ways. For example, what might a student-group-focused dialogic assessment method look like that was highly disruptive to current practices, and aimed at long-term learning?

The first dimension, (1) Temporality, describes when the new assessment method is most relevant. Is the innovation informing exam preparation, the exam situation itself, or does it gain relevance after the exam situation to support evaluation or long-term learning? Does the method and associated tool change over time? The second dimension, Group constellation (2) describes the social arrangement. From a CSCL perspective, the configuration of the group has implications for how collaboration is understood. We might see a group constellation consisting of a single student and single examiner or we might see a dialogic assessment design for groups of students. In Norway, a typical oral exam has a single student with 2 examiners which means that the post-exam evaluation conversation between the two examiners is also a kind of collaboration that could benefit a mediating tool. The Target user (3) describes who has access to the computational artifact in the new assessment method. Whose actions are being mediated? Does the tool support the student, in preparing for an oral exam? Does the tool support the examiner in guiding the assessment dialogue? Is it available to the group?

Table 1
Dimensions for a Given Dialogic Assessment Method

 © ISLS

Dimension	Description	Possibilities
1) Temporality	When in the assessment process is the method relevant?	Preparation → During → Post → Long Term
2) Group constellation	What are the group constellations?	1:1 Student and Examiner → Group
3) Target user	Whose actions are mediated?	Student → Combination → Examiner
4) Interpretation	What levels of interpretation are provided by the tool?	Documentation / Transcription → Analysis / Interpretation
5) Modality	What modalities are introduced?	Hidden Structural → Auditory → Visual → Artifact
6) Ethics	What levels of ethical concern are raised?	Low Risk/ Low Intrusion → High Risk/ High Intrusion
7) Institutional grounding	How transformative is the method in relation to current practices?	Grounded in Current Practices → Highly Disruptive to Current Practices

Another dimension, Interpretation (4) has to do with the kinds of support that are being provided by the computational artifact. Does the tool provide descriptive support, for example in the form of documentation or transcription? Is the tool doing analytic interpretive work in the context of the assessment dialogue? From a CSCL perspective, the last decade has seen growing interest in learning analytics (Wise & Schwartz, 2017) and computational modeling (Baker & Reimann, 2024) that could inform a new assessment method. The Modality (5) dimension characterizes the qualities of the resources and information available in the method. Is the support largely invisible, is there a text or visual interface? Or is the assessment structured around a physical artifact?

Ethical considerations are very important in summative assessment situations, and these can range from issues of privacy to fairness to unforeseen consequences for marginalized groups. The Ethics dimension (6) lifts up these concerns and lets us identify the risks associated with introducing a new assessment method. For example, oral exams can be very stressful situations for students, and it is important the innovations do not magnify these stresses. Finally, it is important to consider the Institutional grounding (7) of the method. How transformative is the method in relation to current practices? New assessment methods that are dramatically different from participant and institutional expectations need to consider the kinds of disruptions that might be introduced to support the adoption of new practices. New methods that build on existing practices may present fewer challenges.

These dimensions, as part of the dialogic assessment design space should provide more structured grounding for future work. They also help us articulate the kinds of collaborative processes which emerge as relevant, the mediational aspects of the computational artifact, and the situatedness of the assessment activity. We can also thus make connections to other CSCL research traditions that might not explicitly deal with summative assessment situations.

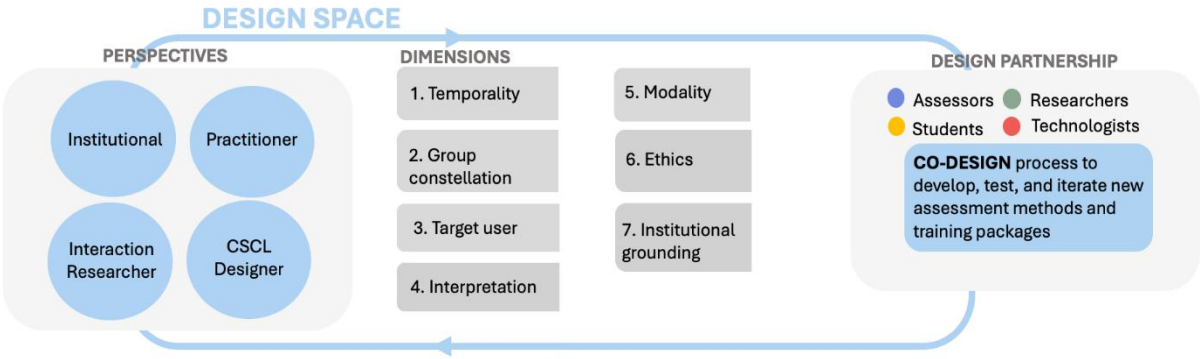
Discussion and concluding remarks

This paper has argued that dialogic assessment formats foster an interactive environment where students actively engage with and apply their knowledge. However, these benefits are contingent on the relational dynamics between students and educators. From a sociocultural and CSCL perspective, oral exams can be understood as assessments of students' individual competence, that are undertaken within collaboratively organized, tool-mediated activities in which knowledge is distributed across students, examiners, and artefacts. We do not claim to resolve this tension, but treat students' individual summative performance as an institutional requirement that needs to be negotiated within a distributed activity system. Our argument is that these interpretive processes, which are already taking place, can be made more transparent and systematic, rather than eliminating the inherently interactional nature of dialogic assessment. For this purpose, we need tools and strategies to mediate what examiners see and support their documentation of student achievement, thereby strengthening the validity of the assessment. For dialogic assessment to be effective, examiners must acknowledge their role in shaping opportunities for students to demonstrate their understanding. This involves being attuned to students' needs while maintaining a clear distinction between the students' contributions and their own (Hontvedt et al., 2024). Another challenge is that dialogic assessment can appear unsystematic with choices being invisible and untraceable. CSCL as a field has much to contribute to these challenges by framing the relevant dialogues as collaborative, and by providing computer-mediated-support to the process. Framing dialogic assessment as a computer-supported and collaborative activity will also bring summative assessment methods into alignment with innovations in learning activity design. This shift calls for support structures that help map emerging possibilities, along with a design space for innovating, refining, and testing methods through co-design in schools and teacher education. The design

space for dialogic assessment (Figure 1) can aid translating overarching objectives into a concrete method that is actionable in concrete subjects and educational settings.

Figure 1

The Design Space for Dialogic Assessment Engages the Four Perspectives in Developing and Refining Concrete Methods Across Seven Dimensions, with The Aim of Creating Actionable Methods to Be Tested in Design Experiments.



The design space provides a shared vocabulary for collaborative development and enables the joint analysis of existing practices while fostering a deeper understanding of the rationale behind previous decisions (Lee et al., 2024). The design space for dialogic assessment builds on a cumulative and multifaceted approach to perspective taking from relevant knowledge areas. We have approached dialogic assessment as highly situated in local contexts and thus began with the relevant institutional perspectives. For us, this means we ground our local approach in the Norwegian education context, but we hope that insights from this perspective can be extended. We then turned to practitioner perspectives to see how examiners and school leaders conceive of oral exams and how they might be strengthened. Next, we introduced perspectives grounded in interaction analysis and video analytic methods. Adopting these lenses in the context of this design space provides insights about how we approach interaction in design work. We thus then turned to a CSCL designer perspective to see what lessons could be learned from engaging in a prototyping process. Here, questions about the relevant technology interfaces could be articulated, as well as questions about what aspects of the assessment process could benefit from designed computer support. Thus, working with a specific prototype created an opportunity to reflect on broader design possibilities. Finally, to create a shared focus across these four perspectives, we developed a set of seven dimensions that ensures progress and structure in design space for dialogic assessment.

In the case of Norway, we find that the design space ensures that assessment configurations are research-based, context-sensitive, and aligned with local goals. By explicating the design space in terms of perspectives and dimensions, the design process becomes transparency, avoiding black-boxing critical elements such as stakeholder involvement and pedagogical principles that underlie certain methods. Having a functioning prototype creates opportunities for conducting design experiments and iteratively refining assessment designs based on empirical results. Regarding generalization, the design space framework can be adapted by other practitioners and researchers to develop dialogic assessment methods that address their specific needs and contexts.

This proposal has several limitations. The empirical basis consists of pilot investigations with a limited number of participants in specific contexts, and we acknowledge that the proposed design space may need to be expanded and revised when confronted with a broader set of relevant actors and contexts. While our empirical investigations are not thorough enough to generalize to the needs and perspectives of a larger group of stakeholders, they do confirm the validity of including these kinds of perspectives in the design space for dialogic assessment and point to their potential importance in future work. As such, this paper provides an initial exploration of the potential to merge learning sciences, CSCL, and interaction analytic insights in order to create theoretically informed design principles and hypotheses about the future of dialogic assessment. Consequently, the design space outlined here should be considered a starting point for collaborative refinement within the CSCL community.

Finally, this partnership-based approach underscores the collaborative nature of the design process, where practitioners, researchers, and technologists work together to create assessments that are both innovative and practical. By expanding on the design space, we aim to bridge the gap between theoretical principles and real-world assessment practices, fostering the sustainable development of dialogic assessment across diverse educational settings.

References

Akimov, A., & Malin, M. (2020). When old becomes new: a case study of oral examination as an online assessment tool. *Assessment & Evaluation in Higher Education*, *45*(8), 1205–1221.

Baker, M., & Reimann, P. (2024). CSCL: a learning and collaboration science?. International Journal of Computer-Supported Collaborative Learning, 19(3), 273-281.

Bayley, T., Maclean, K. D. S., & Weidner, T. (2024). Back to the Future: Implementing Large-Scale Oral Exams. *Management Teaching Review*, *0*(0).

Bernard, R. C., & Kermarrec, G. (2022). Peer assessment and video feedback for fostering self, co, and shared regulation of learning in a higher education language classroom. Frontiers in Education, 7, 1–13.

Boud, D. and Bearman, M. (2022). The assessment challenge of social and collaborative learning in higher education, *Educational Philosophy and Theory*

Chi, M. T. H., Kang, S., & Yaghmourian, D. L. (2017). Why students learn more from dialogue- than monologue-videos: Analyses of peer interactions. The Journal of the Learning Sciences, 26(1), 10–50.

Coelho, R., Logan, C., Agkün, S., & Pea, R. (2025). A Design Space for Articulating and Addressing the Risks of Integrating Generative AI in Education. In *Proceedings of the 19th International Conference of the Learning Sciences-ICLS 2025*, pp. 1679-1683. International Society of the Learning Sciences.

Collins, A., & Halverson, R. (2018). *Rethinking education in the age of technology*. (2nd ed.) Teachers College Press.

Damşa, C. I. (2014). The multi-layered nature of small-group learning: Productive interactions in object-oriented collaboration. *International Journal of Computer-Supported Collaborative Learning, 9*, 247-281.

Daniels, H. (2010). Vygotsky and psychology. In U. Goswami (Ed.), *The Wiley-Blackwell handbook of childhood cognitive development* (pp. 673–696). Wiley-Blackwell.

Davidsen, J., & Steier, R. (2024). Near future practices of interaction analysis: Technology-mediated trends in educational research. International Journal of Research & Method in Education.

Dixon-Woods, M. (2011). Using framework-based synthesis for conducting reviews of qualitative studies. *BMC Medicine*, *9*(1), 39–39. https://doi.org/10.1186/1741-7015-9-39.

Greenblatt, P., & McDonald, P. (2024). Using multimodal public service announcement video projects as communicative language assessments. RELC Journal: A Journal of Language Teaching and Research, 55(2), 559–566.

Hall, R., & Stevens, R. (2015). Interaction analysis approaches to knowledge in use. In *Knowledge and interaction* (pp. 88-124). Routledge.

Hartmann, C. (2025). Oral Exams for a Generative AI World: Managing Concerns and Logistics for Undergraduate Humanities Instruction. *College Teaching*, 1–10.

Hontvedt, M., & Arnseth, H. C. (2013). On the bridge to learn: Analysing the social organization of nautical instruction in a ship simulator. *International Journal of Computer-Supported Collaborative Learning*, *8*(1), 89-112.

Hontvedt, M., Solem, M. S., & Prøitz, T. S. (2024). Oral Exams as Social Practice: Questions, Joint Remembering, and Co-construction. In *Proceedings of the 18th International Conference of the Learning Sciences-ICLS 2024, pp. 402-409*. International Society of the Learning Sciences.

Jordan, B., & Henderson, A. (1995). Interaction analysis: Foundations and practice. *The Journal of the Learning Sciences*, *4*, 39–103.

Lee, M., Gero, K. I., Chung, J. J. Y., Shum, S. B., Raheja, V., Shen, H., ... & Siangliulue, P. (2024). A design space for intelligent and interactive writing assistants. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (pp. 1-35).

Ludvigsen, S., & Steier, R. (2019). Reflections and looking ahead for CSCL: Digital infrastructures, digital tools, and collaborative learning. *International Journal of Computer-Supported Collaborative Learning*, 14(4), 415-423.

MacLean, A., Young, R. M., Bellotti, V. M., & Moran, T. P. (1991). Questions, options, and criteria: Elements of design space analysis. In *Design rationale* (pp. 53-105). CRC Press.

Matuk, C., Tissenbaum, M., & Schneider, B. (2019). Real-time orchestrational technologies in computer-supported collaborative learning: an introduction to the special issue. *International Journal of Computer-Supported Collaborative Learning*, *14*(3), 251-260.

McGrath, C., Farazouli, A., & Cerratto-Pargman, T. (2024). Generative AI chatbots in higher education: a review of an emerging research area. *Higher Education*, 1-17.

Noroozi, O., Teasley, S. D., Biemans, H. J., Weinberger, A., & Mulder, M. (2013). Facilitating learning in multidisciplinary groups with transactive CSCL scripts. *International Journal of Computer-Supported Collaborative Learning*, 8(2), 189-223.

Norwegian Directorate for Education and Training. (2026). *Regler for muntlig eksamen* [Guidelines for oral examinations]. https://www.udir.no/eksamen-og-prover/eksamen/muntlig-eksamen/

Pea, R., Lee, C., Nasir, N. I., & McKinney de Royston, M. (2025). The cultural foundations of learning: Design considerations for measurement and assessment. *Handbook for Assessment in the Service of Learning*, *1*.

Ritchie, S. M. (2016). Self-assessment of video-recorded presentations: Does it improve skills? Active Learning in Higher Education, 17(3), 207–221.

Roza, Z., & Rustam, S. (2023). Digital storytelling to facilitate academic public speaking skills: Case study in culturally diverse multilingual classroom. Journal of Computers in Education, 10, 499–526.

Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education?. *Journal of applied learning and teaching*, *6*(1), 342-363.

Shapiro, B. R., Silvis, D., & Hall, R. (2025). Visualization as theory and experience: interactive qualitative data visualization for the learning sciences. *Journal of the Learning Sciences*, 1–32.

Sidnell, J., & Stivers, T. (Eds.). (2012). *The handbook of conversation analysis*. John Wiley & Sons.

Sikveland, R. O., Solem, M. S., & Skovholt, K. (2021). How teachers use prosody to guide students towards an adequate answer. *Linguistics and Education*, 61.

Skovholt, K., Solem, M. S., Vonen, M. N., Sikveland, R. O., & Stokoe, E. (2021). Asking more than one question in one turn in oral examinations and its impact on examination quality. *Journal of Pragmatics*, *181*, 100–119.

Solem, M. S., Landmark, A. M. D., Stokoe, E., & Skovholt, K. (2023). Assessment in practice: achieving joint decisions in oral examination grading conversations. *Scandinavian Journal of Educational Research*, 1522-1539.

Stahl, G., & Hakkarainen, K. (2021). Theories of CSCL. *International handbook of computer-supported collaborative learning*, 23-43.

Vonen, M. N., Solem, M. S., & Skovholt, K. (2022). Managing students' insufficient answers in oral examinations. *Classroom Discourse*, *14*(3), 258–280.

Acknowledgments

We wish to thank the students, teachers, and colleagues who have participated our workshops, interviews, and informal dialogues about assessment topics. We wish to thank Alexander Stein for developing the first prototype as well as the group of students who provided design reflections with this prototype.

International Society of the Learning Sciences

Flexibility Before Fluency: Exploring How Flexible Strategies Emerge through Collaborative Problem-Solving

Seiyon Lee, University of Florida, leeseiyon@ufl.edu
Shan Zhang, University of Florida, zhangshan@ufl.edu
Hongming Li, University of Florida, hli3@ufl.edu
Ji-Eun Lee, Singapore University of Technology and Design, jieun_lee@sutd.edu.sg
Avery Closser, University of Florida, avery.closser@ufl.edu
Anthony F. Botelho, University of Florida, abotelho@coe.ufl.edu

Abstract: In K-12 math education, it is important not only to foster *fluency* but also *flexibility* in the problem-solving process. Prior works have focused on measuring learning in terms of efficiency or productivity by examining the extent to which students can optimize their problem-solving paths to involve the fewest number of steps. Limited work has paid attention to the notion of flexibility, or the ability to understand and navigate different paths to the same problem. In this study, we examine how middle school students' use a dynamic notation tool to explore flexible solution paths during collaborative problem solving (CPS) activity. Results show that there are differences in patterns of CPS process between students who demonstrated greater flexibility and those who did not. Specifically, they differed in their engagement with higher-level operations. We discuss the implications for supporting students' flexible problem solving and teachers in orchestrating CPS for similar goals in classrooms.

Introduction

An important goal of mathematics learning is to promote both conceptual and procedural knowledge. Students must not only grasp the underlying concepts, but also apply the appropriate strategies. Prior work has examined the extent to which students can optimize this problem-solving process in an efficient (e.g., Newton et al., 2020) or productive manner (e.g., Lee et al., 2022), suggesting that students reach the goal state in the quickest and the most accurate way possible (Hickendorff et al., 2022). Achieving such *fluency*, however, requires students to have developed some level of *flexibility*, or the ability to understand the relative strengths and shortcomings of different strategies before they take the optimal path for a given context (Star & Rittle-Johnson, 2008). While the importance of flexibility has been well-established (e.g., Blöte et al., 2001; Siegler, 1996), little is known about how students demonstrate procedural flexibility while engaging in problem-solving processes.

Over the past decade, several studies have examined instructional approaches to foster flexibility. Some have focused on explicit instruction, directly teaching students multiple strategies (Rittle-Johnson et al., 2012). Others have used worked examples to help students compare and contrast alternative strategies (Newton et al., 2010; Rüede et al., 2023), or encouraged students to generate and evaluate their own solutions (Xu et al., 2017). However, most of this work has focused on individual learning contexts, overlooking how flexibility may emerge during collaborative problem solving (CPS).

In CPS, learners construct a shared cognitive space through social interactions (Miyake & Kirschner, 2014; Roschelle & Teasley, 1995; Swiecki et al., 2020), supporting one another's learning in ways that teachers may not always be able to do in real time. In this regard, the use of computer-supported collaborative learning (CSCL) environments are particularly relevant for facilitating *flexibility* through CPS as they allow students easily to *explore* a wide range of strategies and *reflect* on them in rapid cycles (Mavrikis et al., 2022), which could be difficult in traditional instruction. Yet, few studied how flexibility may be fostered by leveraging such CSCL affordances.

In this study, we investigate how CPS may support the development of flexibility in K–12 mathematics classrooms. Specifically, we orchestrated CPS for 40 sixth-grade middle school students using Graspable Math (GM), where they used dynamic notational tools to coordinate and jointly solve algebra problems on a shared digital whiteboard. We conducted a fine-grained analysis of students' process data, focusing on their use of different algebraic manipulations within GM (e.g., distributing, factoring, canceling terms), perception of collaboration, math beliefs as well as their overall performance. Specifically, our work is guided by the following research questions (RQs):

- RQ1. How do students' flexibility compare when using graded versus binary scoring approaches?
- RQ2. How do students' behavioral indicators of flexibility during CPS relate to their learning?
- RQ3. How do patterns of algebraic actions differ between higher- and lower-performing groups?

 © ISLS

Related Work

Procedural flexibility through CPS

The development of flexibility is a critical yet challenging goal in K–12 mathematics classrooms. Research has shown that students who are flexible problem-solvers tend to have deeper conceptual knowledge and exhibit greater fluency in the problem-solving process (Blöte et al., 2001; Siegler, 1996). While teachers can provide instruction, the effectiveness of the intervention significantly varies within and across contexts, and importantly students often revert to a single, standard solution (Star et al., 2022). Prior work has primarily focused on the specific types of instruction that teachers can provide; however, even well-designed curricula offer limited opportunities for students to explore a range of strategies and solution paths (Hickendorff & van Zanten, 2024). Less attention has been given to the role that students themselves can play in the process (Munaji et al., 2025).

One of such under-explored possibilities is CPS, a socio-cognitive process in which learners take turns on the same task and build a shared understanding, which in turn shapes their subsequent interactions (Clark, 1996). CPS is recognized in K-12 math education as an opportunity for students to articulate mathematical ideas and reasoning (NCTM, 2024; Moses & Cobb, 2001; Mullins et al., 2011;). Studies also found that flexibility is improved when students engage in group work and math discussions (Durkin et al., 2021).

While a substantial body of research has examined CPS, most studies examined students' discourse or other coarse-grained artifacts (e.g., Hu, 2021). These approaches are often limited, either by students' ability to verbalize or demonstrate their knowledge, or by the time and resources required for researchers to qualitatively code such data. Therefore, there is a need for novel tools that not only facilitate CPS activity, but also capture the process at a more granular level, for instance through the automatic tracking and tagging of students' behaviors in the CSCL environment.

Affordances of dynamic notation tools

The emergence of dynamic notation tools helped provide a much-needed affordance. Often embedded in digital whiteboards (e.g., MathSpace, FluidityMath, Graspable Math), students use these tools to interact with numbers and mathematical symbols through specified gesture-actions (e.g., dragging, and tapping). As they are largely analogous to the real dynamics of solving algebra problems, students can focus on identifying the visual cues in the mathematical notations, grasping their meaning, and navigating different variations in the equation (Lindgren & Johnson-Glenberg, 2013; Ottmar et al., 2015). Studies have shown that embodied experiences shape math thinking, and improve learning and engagement (Abrahamson et al., 2020; Goldstone et al., 2017; Landy & Goldstone, 2010; Ottmar & Landy, 2017; Weitnauer et al., 2016).

Dynamic notation tools are well-suited for studying flexibility or CPS more broadly for two reasons. First, they allow students to explore multiple approaches with minimal scaffolding to guide the process. For example, students' actions are immediately followed by real-time visual transformations that provide feedback on their operations. When students attempt to perform an invalid operation, the system simply does not execute it, often signaling this through subtle visual cues or responsive motion. This design helps avoid growing any misconceptions while reinforcing underlying mathematical principles. Combined with interactive features, these affordances enable rapid, low-risk experimentation from which students can learn to discern which strategies may be more appropriate under certain conditions (Komatsu & Jones, 2020). Importantly, all of this process is automatically recorded and stored in the form of log data, which can later be used for post-hoc analyses (Chan et al., 2023). Recent studies have begun using such process data to unpack CPS dynamics, though most have remained descriptive in nature (e.g., Li et al., 2024; Ma et al., 2023).

We leverage one of such tools in *Graspable Math* (GM), where students can encounter and discover algebraic principles by physically moving and rearranging numbers and symbols in a digital whiteboard, much like a mathematical playground. Their interactions are automatically logged as either algebraic or non-algebraic, or as sub-operations of algebraic actions (e.g., *commuting terms*, *canceling fractions*). Prior studies have shown that a gamified version of GM can enhance algebra learning outcomes (Chan et al., 2022; Decker-Woodrow et al., 2023). In this study, we explore the potential of GM to naturally create opportunities for cultivating flexibility through CPS activities.

Methodological approaches to modelling CPS

The CPS process is often complex and difficult to represent in ways that translate into actionable insights for teachers (e.g., determining whether students successfully executed a set of strategies). To model CPS effectively, it is important to account for how each event at a given moment is influenced by what has immediately preceded it or by recent contextual interactions (Ruis et al., 2019). Several methods have been used to reflect the temporal and interdependent nature of CPS processes, such as analyzing the co-occurrence of codes (Shaffer et al., 2017)

 © ISLS

or the sequential order of events (Fogel et al., 2021; Swiecki et al., 2019). However, when used in isolation, these methodological approaches risk either overlooking the temporal dynamics of CPS or over-representing micro-sequences that may not be meaningful.

In this sense, Ordered Network Analysis (ONA; Tan et al., 2022) is a promising approach that accounts for the interactive, interdependent, and temporal aspects of CPS. ONA involves constructing *directed* network models, which visualize how patterns of interactions unfold over time. These models are relatively easy and intuitive to interpret and support both visual and statistical comparisons. For example, individual networks can be combined and compared across groups, such as between higher- and lower-performing students. ONA visualizations can also be incorporated into broader CSCL tools that help teachers monitor how students engage in CPS, as partially demonstrated in prior work (Echeverria et al., 2025). In this study, we use ONA for its dual benefits of modelling students' CPS processes in a way that also holds the potential to inform the future design of teacher-facing orchestration tools.

Methods

Study context and procedure

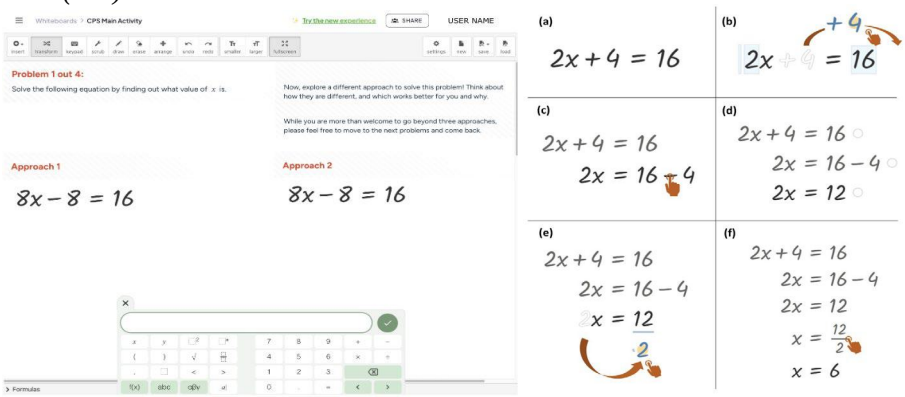
The current study used a recent prototype in Graspable Math (GM), a collaborative whiteboard that enables student pairs to work synchronously on shared problems within a common workspace. Students were seated face-to-face and coordinated their activity through the shared whiteboard, using drag-and-drop gestures to manipulate the digital mathematical representations. This design supports joint engagement, with both students contributing through shared interaction rather than individual control.

Data were collected in a one-week classroom-based study with 43 sixth-grade students from two class periods at a developmental research school in the southern United States. The study was embedded in regular mathematics instruction and conducted over four consecutive days: Day 1 onboarding and a pre-survey on group work perceptions and mathematics beliefs; Day 2 an individual pre-test; Day 3 random pairing for collaborative problem-solving on four problems (Figure 1); and Day 4 an individual post-test and post-survey. All tasks and assessments were conducted in Graspable Math (GM), and surveys were administered via Qualtrics. The study was approved by the local Institutional Review Board.

For analysis, we retained 28 students (14 pairs) after excluding those (and their partners) absent from any pre-test, post-test, or CPS activity. With one missing pre-survey, the final sample included 14 female, 11 male, and 2 non-binary students; 45.8% White, 20.8% Hispanic/Latinx, 16.7% Black, 8.3% Asian, and 8.4% multiracial or other.

Figure 1

Example Problem Layout for CPS activity and Types of Manipulations Possible Through Graspable Math (GM)



Note. For this study, pairs of students solved a set of algebraic equations on the digital whiteboard (left) using gesture-based actions (right; a–f). These manipulations included (b, e) moving variables, coefficients, and constants; (c) tapping to perform basic arithmetic or cancel fractions; and (d) using hidden circles to reposition entire equations.

Measures and data preprocessing

Given the novelty of the study focus, the survey items were informed by and adapted from prior works, such as K-12 mathematics education (e.g., Duran et al., 2014), collaborative learning (e.g., Dewiyanti et al., 2007), and

engagement in digital environments (e.g., Xu & Correia, 2024). The pre-survey included 12 items measuring groupwork attitudes and 7 items assessing mathematics beliefs. The post-survey included 21 items focused on students' perceptions of the collaborative activity, including open-ended questions (e.g., "*Can you describe how you and your partner worked together to discover or explore a different approach to solving the problem*?"). All items measured on a 5-point Likert-scale. For analysis, composite scores were computed by taking the average from the items within each construct.

Students' flexible problem solving was evaluated through a pre-test, a CPS activity, and a post-test. All assessments were completed within the same Graspable Math (GM) environment to ensure consistency across conditions. Each assessment included three algebraic equation tasks and one ratio task. For this set of problems, students were encouraged to elicit at least two possible solution strategies. For the purposes of this study, we focused on the three algebra tasks, as they were most relevant for examining flexibility. Consistent with the survey measures, performance scores were averaged across the three items. Differences between the pre- and post-tests, as well as between the CPS activity and the post-test, were also computed.

As an alternative indicator of flexibility, we also considered students' behavioral variability during the CPS activity. Specifically, three variance measures were derived: (1) *Variance in Action Types* (VAT), capturing variability in the distribution of algebraic and non-algebraic manipulations (e.g., editing texts); (2) *Variance in Algebraic Actions* (VAA), capturing variability in the types of algebraic operations performed; and (3) *Variance in Key Strategies* (VKS), capturing variability in the use of algebraic strategies that are most relevant to the task (e.g., *AddSubInvert*, *MulDivInvert*, *FractionCancelTerms*; see full list in Table 1). Higher values from these measures indicate greater diversity in students' use of available manipulations, algebraic operations, or key strategies,and vice versa. All of these action and strategy types were automatically derived from GM log data, as the system classifies each user action into predefined categories or codes.

Table 1

Types of Algebraic Actions that are Most Relevant to the CPS Task

Algebraic Action	N (%)	Description
AddSubInvert	2,651 (15.2%)	Move the constant to the other side of the equation. (e.g., $8x-8=16 \rightarrow 8x-16-8=0$)
MulDivInvert	2,391 (13.7%)	Invert both sides to simplify expressions into fractions. (e.g., $8x = 24 \rightarrow x = \frac{24}{8}$)
Commute	2,118 (12.2%)	Reorder terms in an equation to change their sequence. (e.g., $x8-24=0 \rightarrow -24+x8=0$)
FractionCancelTerms	1,289 (7.4%)	Cancel fractions to convert expressions into integers. (e.g., $x = \frac{24}{8} \rightarrow x = 3$)
AddSubNumbers	1,267 (7.3%)	Combine or simplify constant terms within the equation. (e.g., $-16-8+8x=0 \rightarrow -24+8x=0$)
DistributeIntoBrackets	1,139 (6.5%)	Apply the distributive property to remove brackets. (e.g., $8\left(x-3\right)=0 \rightarrow x8-24=0$)
FactorCommonTerms	1,010 (5.8%)	Factor out the common term shared across expressions. (e.g., $8x-8=16 \rightarrow 8(x-1)=16$)
GreatestCommonFactor	773 (4.4%)	Factor out the greatest common factor. (e.g., $-16+8x-8=0 \rightarrow (-2+x)8-8=0$)
Total	12,638 (100%)	--

Analytic approach

Data analyses addressed the three RQs using complementary quantitative methods. To address RQ1, we examined whether students' flexibility improved after the CPS activity. Specifically, we used two scoring approaches. The first was a graded scale (0–2) that indicated the number of distinct correct solutions students generated (0 = no correct solution, 1 = one correct solution, 2 = two distinct correct solutions). Because the transition from producing one correct solution to generating two represents a qualitatively greater cognitive shift, we also employed a binary measure that distinguished between students who generated two distinct solutions and those who did not (1 = flexible, 0 = not flexible). This binary measure captures conceptual gains that might be underestimated by a linear scale assuming incremental improvement. Differences between pre- and post-tests, as well as between the CPS activity and the post-test, were then computed. Prior to the analyses, the normality of difference scores was assessed using the Shapiro–Wilk test. Based on the results, the appropriate parametric (paired-samples t-test) or

 © ISLS

non-parametric (Wilcoxon signed-rank test) was applied. Cohen's d was also calculated to estimate effect sizes, and statistical significance was evaluated at $\alpha = .05$. Afterwards, a graded scale was used for the learning measure in subsequent analyses.

To address RQ2, we conducted Pearson's correlation to examine the relationships among students' learning (e.g., pre-, CPS activity, post- scores), survey measures (i.e. math beliefs, groupwork attitudes, and perceptions of Collaboration), and indicators of behavioral variability (e.g., VAT, VAA, and VKS, described above). Then, Benjamini–Hochberg (BH) correction was applied to control the false discovery rate at $\alpha = .05$.

To address RQ3, we conducted ONA to examine patterns of students' algebraic strategy use during the CPS activity. The ONA algorithm identifies and accumulates connections between the current line of data (in this case, a specific strategy students demonstrated) and preceding lines within a defined window. ONA is particularly useful in contexts where the precise sequence of actions is less important than their temporal overlap, such as when students deviate from a prescribed order of operations. The moving window (or stanza) length was set to 14, meaning that actions within this range were considered temporally related. Alternative window sizes had been tested, and the final selection yielded the most moderate goodness of fit.

ONA was implemented with individual students as the unit of analysis, generating a separate network graph for each participant's sequence of actions within their collaborative whiteboard. This resulted in 28 individual-level networks across two groups (Improved vs Not Improved) based on learning gains (i.e., the difference between pre- and post-test scores). Students with positive gain values were classified as the higher-performing group, comprising 15 of the 28 students, while the remaining 13 students were classified as lower-performing. The overall sample had a mean learning gain of 0.44 (SD = 0.72).

Results

RQ1. Effect of CPS on flexibility

We compared students' pre- and post-test flexibility scores using both the graded scale and binary measures as described above. A Shapiro–Wilk test indicated that the graded scores did not significantly deviate from normality ($W = 0.94$, $p = .051$), whereas the binary scores showed a significant deviation ($W = 0.59$, $p < .001$). Accordingly, both parametric (*t*-test) and non-parametric (Wilcoxon *W* signed-rank) analyses were conducted to ensure robustness.

For the graded scale, both analyses revealed a significant improvement from pre- to post-test, $t(n–1) = 3.09$, $p = .004$; $W = 42.50$, $p = .004$, with a medium effect size ($d = 0.54$). In contrast, for the binary measure of flexibility (coded as 1 for flexible and 0 otherwise), the t-test suggested a marginally significant increase, $t(n–1) = 1.97$, $p = .058$, and the Wilcoxon test yielded a similar trend, $W = 4.00$, $p = .059$, with a small-to-medium effect size ($d = 0.34$). This finding is consistent with our hypothesis that it is more difficult for students to come up with a second/alternative approach to solve the same problem.

Overall, results suggest that students became more flexible in their problem-solving after engaging in the CPS activity. Improvements were most evident when flexibility was measured by the linear scale, indicating that many students were able to apply multiple solutions more effectively after CPS. The smaller, marginal change in the binary measure aligns with the hypothesis that producing two distinctly different solutions requires a higher level of cognitive demand.

RQ2. Correlations within CPS processes

In the correlation analysis, three associations remained statistically significant after Benjamini–Hochberg (BH) correction (see Table 2). The first was a strong negative correlation between pre-test score and gain from the pre-test to the CPS activity ($r = -0.76$, $p < .001$). This finding indicates that students with lower initial flexibility tended to show greater relative improvement during collaboration. Similarly, pre-test score was negatively correlated with overall learning gain ($r = -0.69$, $p < .01$), reflecting the same pattern across a broader timeframe. Finally, overall learning gain was positively correlated with gain during the CPS activity ($r = 0.69$, $p < .01$), suggesting that students who showed greater improvement during CPS also tended to exhibit larger overall gains. These relationships suggest that changes in flexibility were related to students' starting points and that progress observed during CPS was closely aligned with cumulative learning across the intervention.

Several associations emerged prior to applying the BH correction, which provides additional insight into potential relationships among behavioral and learning measures. Variance in the types of actions (including both algebraic and non-algebraic manipulations) was positively associated with CPS activity score ($r = .45$, $p < .05$) and with learning gain between pre-test and the CPS activity ($r = .52$, $p < .01$). These findings suggest that students who engaged in a wider range of interaction types tended to perform better during collaboration and showed

 © ISLS

greater improvement from their initial levels of flexibility. Similarly, variance in algebraic actions was positively correlated with post-test performance (r = .41, p < .05), indicating that students who exhibited greater variability in their algebraic operations also tended to perform better individually after collaboration. Considering that students engaged in only one day of CPS activity, this pattern supports the idea that behavioral diversity, reflected in variability across action types, may serve as an indicator of a flexible problem-solving process during CPS. Importantly, the diversity of interactions appeared to matter more than whether students explored specific algebraic strategies or whether those actions were directly relevant to the task.

It is also worth noting that students' math beliefs were positively correlated with both their perceptions of collaboration (r = .49, p < .01) and their post-test performance (r = .41, p < .01), highlighting a potential link between positive mathematical dispositions, collaborative attitudes, and achievement. Moreover, post-test scores were positively associated with CPS activity performance (r = .40, p < .01), and CPS activity performance itself was strongly correlated with immediate learning gain from pre-test to CPS activity (r = .69, p < .01). Overall, these patterns suggest that students who engaged more flexibly and held stronger beliefs about mathematics and collaboration tended to achieve higher levels of performance both during and after the CPS activity.

Table 2

Correlations Among CPS Measures

Measures	1	2	3	4	5	6	7	8	9	10
1 Variance in Action Types	-									
2 Variance in Algebra Actions	.35	-								
3 Variance in Key Strategies	−.13	-.05	-							
4 Math Belief	-.07	-.14	.12	-						
5 Perceptions of Collaboration	-.10	-.21	.03	**.49**	-					
6 Pre-Test Score	-.25	.01	.07	.17	.24	-				
7 CPS Activity Score	**.45**	.32	.03	.25	.09	.21	-			
8 Post-Test Score	.07	**.41**	-.02	**.41**	.14	.36	**.40**	-		
9 Learning Gain (Post – Pre)	.29	.31	-.08	.16	-.12	**-.69**	.11	.43	-	
10 CPS Activity – Pre	**.52**	.20	-.04	.01	-.15	**-.76**	**.48**	-.06	**.69**	-

Note. Significant correlations are in bold; bh-corrected ones are highlighted in grey.

RQ3. CPS patterns between groups

To better understand how students engaged in CPS at a finer-grained level, we compared the aggregated mean ONA networks between groups of students who showed improvement in flexibility and those who did not. (Figure 2). In ONA, the node size is proportional to the number of occurrences of that code as a response to other codes in the data, while the color and saturation of the circle within each node are proportional to the number of self-connections for that code. In this regard, we found that the similar sets of codes (or algebraic actions) were large in both groups. For instance, both groups responded most frequently with *MulDivInvert* and *Other*, which were followed by *AddSubInvert*. These reflect codes with more frequent self-connections, or where a code is both what students responded to and what they responded with.

At a high level, the high-performing group exhibited fewer salient connections (with line weight LW > 0.1) compared to the low-performing group. This suggests that successful students engaged in a more strategic process, relying on a smaller set of carefully sequenced moves rather than exploring a wider variety of actions. In contrast, the low-performing group's more complex network of transitions indicates exploratory behavior, which may be less consistent in terms of control. These findings imply that effective flexibility in algebraic problem solving may depend less on the number of strategies employed and more on the purposeful coordination of strategies in particular sequences.

More importantly, these groups also exhibited notably distinct patterns in terms of the directionality between key algebraic actions. Given the ways tasks are designed, certain sequences of algebra actions are more logical and sensible than others. For instance, if Invert both sides to simplify expressions into fractions remaining x in one side and constant in others (MulDivInvert) and then it naturally follows Cancel fractions to convert expressions into integers (FractionCancelTerms).

Results showed that patterns of key algebraic actions mostly overlapped between the higher- and lower-performing groups. For example, both groups demonstrated transitions such as *AddSubInvert* → *MulDivInvert*, *AddSubNumbers* → *MulDivInvert*, and *MulDivInvert* → *FractionCancelTerm*. This suggests that students engaged

 © ISLS

in lower-level alternation, such as between addition–subtraction and multiplication–division operations, which are fundamental to the process of systematically restructuring the algebraic equation.

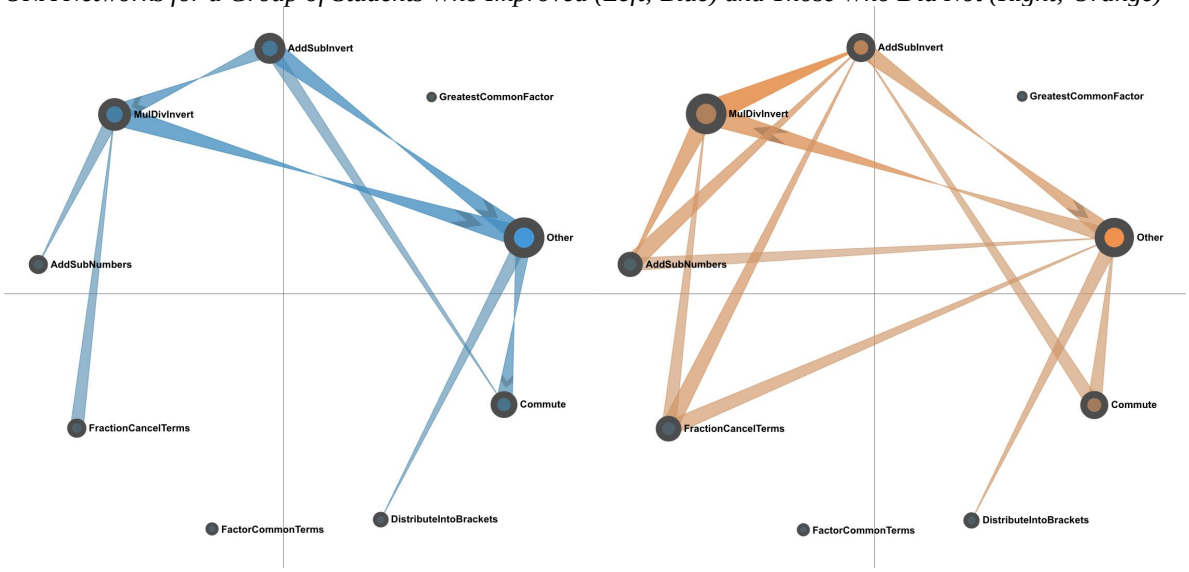
However, the two groups differed in their involvement of *Other*, which represents algebraic actions less directly relevant to solving the given problem. The higher-performing group showed more frequent transitions from *MulDivInvert → Other* (LW = .17), indicating that these students often engaged in higher-level operations before exploring alternative approaches. In contrast, the lower-performing group exhibited the opposite transition, *Other → MulDivInvert* (LW = .16), suggesting a possible reliance on inversion operations that could complicate the equations, particularly when executed before higher-level transformations had been applied.

As mentioned earlier, a wider range of action sequences was observed for the lower-performing group. For example, the higher-performing group displayed unique connections from *Commute → AddSubInvert*, which were absent in the lower-performing group. In contrast, the lower-performing group demonstrated transitions such as *AddSubInvert → AddSubNumbers*, *AddSubInvert → FractionCancelTerms*, and *Other → FractionCancelTerms*. These patterns suggest a tendency to move quickly among related but less coordinated operations.

Overall, the broader and less focused set of connections in the lower-performing group points to a more exploratory, trial-and-error approach to problem solving, whereas the higher-performing group exhibited more deliberate and strategically organized transitions among the same set of algebraic operations.

Figure 2

ONA Networks for a Group of Students Who Improved (Left; Blue) and Those Who Did Not (Right; Orange)



Note. Only the most salient connections (line weights LW > 0.1) are displayed to facilitate interpretability.

Table 3

Most Salient Patterns between High- and Low- Performing Groups (with Line Weight of 0.10 or Above)

No	High-Performing Group			Low-Performing Group		
	From	To	LW	From	To	LW
1	**MulDivInvert**	**Other**	**0.17**	AddSubInvert	MulDivInvert	0.20
2	AddSubInvert	Other	0.16	**Other**	**MulDivInvert**	**0.16**
3	AddSubInvert	MulDivInvert	0.15	AddSubNumbers	MulDivInvert	0.15
4	Other	Commute	0.14	AddSubInvert	Other	0.14
5	AddSubNumbers	MulDivInvert	0.11	**AddSubInvert**	**AddSubNumbers**	**0.14**
6	MulDivInvert	FractionCancelTerms	0.11	**AddSubInvert**	**FractionCancelTerms**	**0.13**
	Commute	**AddSubInvert**	**0.11**	Distrib...Bracket	Other	0.12
7	Distrib...Bracket	Other	0.11	**Other**	**FractionCancelTerms**	**0.12**
8	-	-	-	MulDivInvert	FractionCancelTerms	0.11
9	-	-	-	Other	Commute	0.11

Note. Contrasting patterns between two groups are in bold.

Discussion

This study examined how students' flexibility unfolds during CPS in algebra. Results showed that participation in the CPS activity was associated with improvements in mathematical flexibility, particularly among students who began with lower initial knowledge. Although variance measures and mathematics beliefs did not reach statistical significance within this single instance of CPS, the findings suggest that these factors may still play a meaningful role. Furthermore, the observed gains appeared to depend less on the number of operations performed than on the purposeful coordination of strategies that distinguished higher- from lower-performing students. This pattern aligns with prior research emphasizing that flexible problem solvers are not only capable of executing multiple strategies but can also select and apply them adaptively based on problem structure (Heinze et al., 2009; Krems, 2014; Star & Rittle-Johnson, 2008).

Moreover, high-performing students in this study appeared to engage in high-level thinking, monitoring and regulation of their problem-solving processes, using the digital whiteboard environment to test and refine different strategies, as opposed to playing around with drag-and-drop interactions. In contrast, low-performing students exhibited more reactive and locally driven transitions, either reflecting procedural rigidity or lacking deliberate coordination among algebraic operations. One possible explanation for these differences may relate to affective factors such as mathematics anxiety, which has been shown to impede reasoning and reduce students' willingness to explore alternative strategies despite challenges or unexpected turns of events (Jiang et al., 2020). These differences highlight a competence–performance gap well-documented in the literature in that knowing multiple strategies does not necessarily entail using them flexibly in practice (Hickendorff, 2022).

The observed patterns highlight the importance of designing adaptive supports that can help teachers recognize and respond to students' differing levels of flexibility in real time. Future work could explore the integration of orchestration tools or dashboards that visualize students' problem-solving trajectories and alert teachers to moments of unproductive repetition or disengagement. Such tools could enable timely interventions, particularly for students who may struggle to coordinate strategies or persist through uncertainty.

Limitations and future directions

Several limitations should be noted, which can inform directions for future research. First, this study focused on the design and development of three sets of algebra problems that built upon one another while introducing alternative ways to solve equations. Although the tasks explicitly prompted students to generate alternative approaches, the representational formats used in this study may have constrained the range of strategies students could explore. Relatedly, prior research has shown that the presentation of problems can influence students' strategy use and flexibility (Chan et al., 2022), suggesting that our findings may be somewhat conditional on the representational formats employed in this study. Future work can broaden the range of problem representations so that students are encouraged to explore different strategies in more flexible ways.

Second, this study was limited to observable actions as recorded by the platform. While this feature makes the data well suited for developing tools that can be used in real time, it does not capture students' cognitive and metacognitive processes, which are often latent and invisible but may play a critical role in shaping how students strategically engage in CPS—for example, through self-correction or reflection between moves. In this regard, prior work has combined transmodal analysis with ONA to provide a more contextualized understanding of learning processes (e.g., Borchers et al., 2024). However, further research is needed to examine how students coordinate cognitive and embodied aspects of problem solving within such integrated frameworks.

Last, this study did not examine individual-level ONA networks. Aggregating networks at the group level may have obscured within-group variation in how CPS differentially supported or hindered the development of procedural flexibility. Future research could analyze unit- or participant-level differences to explore how specific peer interactions contribute to individual learning trajectories and the development of flexibility over time.

Conclusion

This study highlights flexibility as a distinct dimension of mathematical problem solving, complementing traditional measures of fluency and efficiency. By examining students' interactions with a dynamic notation tool in a CSCL environment, we found that greater flexibility was linked to higher-level algebraic reasoning during CPS. These findings demonstrate the potential of such tools to both reveal and foster flexible problem-solving processes in K–12 mathematics classrooms. To better support flexibility across different levels of proficiency, it is important to design opportunities for students to explore multiple strategies with appropriate structure and guidance. Equipping teachers with tools to monitor and provide timely, targeted interventions can further promote equitable and inclusive opportunities for all learners to engage in meaningful mathematical exploration.

 © ISLS

References

Abrahamson, D., Nathan, M. J., Williams-Pierce, C., Walkington, C., Ottmar, E. R., Soto, H., & Alibali, M. W. (2020, August). The future of embodied design for mathematics teaching and learning. In *Frontiers in Education* (Vol. 5, p. 147). Frontiers Media SA.

Blöte, A. W., Van der Burg, E., & Klein, A. S. (2001). Students' flexibility in solving two-digit addition and subtraction problems: Instruction effects. *Journal of educational psychology*, *93*(3), 627.

Borchers, C., Wang, Y., Karumbaiah, S., Ashiq, M., Shaffer, D. W., & Aleven, V. (2024, March). Revealing networks: Understanding effective teacher practices in AI-supported classrooms using transmodal ordered network analysis. In *Proceedings of the 14th Learning Analytics and Knowledge Conference* (pp. 371-381)

Chan, J. Y. C., Closser, A. H., Smith, H., Lee, J. E., Drzewiecki, K. C., & Ottmar, E. (2023). Grasping patterns of algebraic understanding: Dynamic technology facilitates learning, research, and teaching in mathematics education. In *Mathematical Teaching and Learning: Perspectives on Mathematical Minds in the Elementary and Middle School Years* (pp. 207-225). Cham: Springer International Publishing.

Chan, J. Y. C., Ottmar, E. R., Smith, H., & Closser, A. H. (2022). Variables versus numbers: Effects of symbols and algebraic knowledge on students' problem-solving strategies. *Contemporary Educational Psychology*, *71*, 102114.

Chan, J. Y. C., Lee, J. E., Mason, C. A., Sawrey, K., & Ottmar, E. (2022). From Here to There! A dynamic algebraic notation system improves understanding of equivalence in middle-school students. *Journal of Educational Psychology*, *114*(1), 56.

Clark, H. H. (1996). *Using language*. Cambridge university press.

Decker-Woodrow, L. E., Mason, C. A., Lee, J. E., Chan, J. Y. C., Sales, A., Liu, A., & Tu, S. (2023). The impacts of three educational technologies on algebraic understanding in the context of COVID-19. *AERA open*, *9*, 23328584231165919.

Durkin, K., Rittle-Johnson, B., Star, J. R., & Loehr, A. (2023). Comparing and discussing multiple strategies: An approach to improving algebra instruction. *The Journal of Experimental Education*, *91*(1), 1-19.

Echeverria, V., Zhao, L., Alfredo, R., Milesi, M. E., Jin, Y., Abel, S., ... & Martinez-Maldonado, R. (2025, April). TeamVision: An AI-powered Learning Analytics System for Supporting Reflection in Team-based Healthcare Simulation. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (pp. 1-22).

Fogel, A., Swiecki, Z., Marquart, C., Cai, Z., Wang, Y., Brohinsky, J., ... & Shaffer, D. W. (2021, January). Directed epistemic network analysis. In *International Conference on Quantitative Ethnography* (pp. 122-136). Cham: Springer International Publishing.

Goldstone, R. L., Marghetis, T., Weitnauer, E., Ottmar, E. R., & Landy, D. (2017). Adapting perception, action, and technology for mathematical reasoning. *Current Directions in Psychological Science*, *26*(5), 434-441.

Heinze, A., Star, J.R. & Verschaffel, L. Flexible and adaptive use of strategies and representations in mathematics education. *ZDM Mathematics Education* 41, 535–540 (2009).

Hickendorff, M., & van Zanten, M. (2024). Opportunity to learn flexible and adaptive strategy use in current and past Dutch mathematics textbooks. *Mathematical Thinking and Learning*, 1-18.

Hickendorff, M., McMullen, J., & Verschaffel, L. (2022). Mathematical flexibility: Theoretical, methodological, and educational considerations. *Journal of Numerical Cognition*, *8*(3), 326-334.

Hickendorff, M. (2022). Flexibility and adaptivity in arithmetic strategy use: What children know and what they show. *Journal of Numerical Cognition*, *8*(3), 367-381.

Hu, L. (2021). Turn-usurping in dialogic collaborative problem solving. In *Proceedings of the 15th International Conference of the Learning Sciences-ICLS 2021*.. International Society of the Learning Sciences.

Jiang, R., Liu, R. D., Star, J., Zhen, R., Wang, J., Hong, W., ... & Fu, X. (2021). How mathematics anxiety affects students' inflexible perseverance in mathematics problem-solving: Examining the mediating role of cognitive reflection. *British Journal of Educational Psychology*, *91*(1), 237-260.

Komatsu, K., & Jones, K. (2020). Interplay between paper-and-pencil activity and dynamic-geometry-environment use during generalisation and proving. *Digital Experiences in Mathematics Education*, *6*(2), 123-143.

Krems, J. F. (2014). Cognitive flexibility and complex problem solving. In Complex problem solving (pp. 201-218). Psychology Press.

Landy, D., & Goldstone, R. L. (2010). Proximity and precedence in arithmetic. *Quarterly journal of experimental psychology*, *63*(10), 1953-1968.

Li, H., Zhang, S., Lee, S., Lee, J. E., Zhong, Z., Weitnauer, E., & Botelho, A. F. (2024). Math in motion: Analyzing real-time student collaboration in computer-supported learning environments. In *Proceedings of the 17th International Conference on Educational Data Mining* (pp. 533-541).

Lindgren, R., & Johnson-Glenberg, M. (2013). Emboldened by embodiment: Six precepts for research on embodied learning and mixed reality. *Educational researcher*, *42*(8), 445-452.

Loes, C. N., & Pascarella, E. T. (2017). Collaborative learning and critical thinking: Testing the link. *The Journal of Higher Education*, *88*(5), 726-753

Ma, Y., Zhang, H., Ni, L., & Zhou, D. (2023). Identifying collaborative problem-solver profiles based on collaborative processing time, actions and skills on a computer-based task. *International Journal of Computer-Supported Collaborative Learning*, *18*(4), 465-488.

Mavrikis, M., Rummel, N., Wiedmann, M., Loibl, K., & Holmes, W. (2022). Combining exploratory learning with structured practice educational technologies to foster both conceptual and procedural fractions knowledge. *Educational technology research and development*, *70*(3), 691-712.

Melcer, E. F., & Isbister, K. (2016, May). Bridging the physical divide: a design framework for embodied learning games and simulations. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 2225-2233).

Miyake, N., & Kirschner, P. A. (2014). The social and interactive dimensions of collaborative learning.

Munaji, M., Rohaeti, T., Mutadi, M., Sumliyah, S., & Kodirun, K. (2025). A literature review of flexibility in interactive mathematics classrooms: the role of teachers and students. Journal of Education and Learning (EduLearn), 19(2), 597-605.

National Council of Teachers of Mathematics (2024). Principles to Actions: Ensuring Mathematical Success for All. 10th Anniversary Edition. NCTM.

Newton, K. J., Lange, K., & Booth, J. L. (2020). Mathematical flexibility: Aspects of a continuum and the role of prior knowledge. *The Journal of Experimental Education*, *88*(4), 503-515.

Ottmar, E., & Landy, D. (2017). Concreteness fading of algebraic instruction: Effects on learning. *Journal of the Learning Sciences*, *26*(1), 51-78.

Ottmar, E. R., Landy, D., Goldstone, R., & Weitnauer, E. (2015). Getting From Here to There!: Testing the effectiveness of an interactive mathematics intervention embedding perceptual learning. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 37).

Rittle-Johnson, B., & Star, J. R. (2009). Compared with what? The effects of different comparisons on conceptual knowledge and procedural flexibility for equation solving. *Journal of Educational Psychology*, *101*(3), 529.

Roschelle, J., & Teasley, S. D. (1995, August). The construction of shared knowledge in collaborative problem solving. In *Computer supported collaborative learning* (pp. 69-97). Berlin, Heidelberg: Springer Berlin Heidelberg.

Ruis, A., Siebert-Evenstone, A., Pozen, R., Eagan, B., & Shaffer, D. W. (2019). Finding common ground: A method for measuring recent temporal context in analyses of complex, collaborative thinking.

Star, J. R., & Newton, K. J. (2009). The nature and development of experts' strategy flexibility for solving equations. *ZDM*, *41*(5), 557-567.

Swiecki, Z., Lian, Z., Ruis, A., & Shaffer, D. W. (2019). Does order matter? Investigating sequential and cotemporal models of collaboration.

Swiecki, Z., Ruis, A. R., Farrell, C., & Shaffer, D. W. (2020). Assessing individual contributions to collaborative problem solving: A network analysis approach. *Computers in Human Behavior*, *104*, 105876.

Tan, Y., Ruis, A. R., Marquart, C., Cai, Z., Knowles, M. A., & Shaffer, D. W. (2022, October). Ordered network analysis. In *International Conference on Quantitative Ethnography* (pp. 101-116). Cham: Springer Nature Switzerland.

Tan, Y., Swiecki, Z., Ruis, A. R., & Shaffer, D. (2024). Epistemic network analysis and ordered network analysis in learning analytics. In *Learning Analytics Methods and Tutorials: A Practical Guide Using R* (pp. 569-636). Cham: Springer Nature Switzerland.

Weitnauer, E., Landy, D., & Ottmar, E. (2016, December). Graspable math: Towards dynamic algebra notations that support learners better than paper. In *2016 Future Technologies Conference (FTC)* (pp. 406-414). IEEE.

International Society of the Learning Sciences

Using Multimodal Analytics-Supported Reflective Assessment to Promote Undergraduates' Productive Social-emotional Interaction

Yuqin Yang, Central China Normal University, China, yangyuqin@ccnu.edu.cn
Xiaoru Li, Central China Normal University, China, XiaoRu2002@mails.ccnu.edu.cn
Daner Sun, The Education University of Hong Kong, Hong Kong S.A.R., dsun@eduhk.hk
Xueqi Feng, Southern University of Science and Technology, China, fengxq@sustech.edu.cn
Gaoxia Zhu, Nanyang Technological University, Singapore, gaoxia.zhu@nie.edu.sg

Abstract: This quasi-experimental study investigated the effects of multimodal analytics-supported reflective assessment (MASRA) on undergraduates' collaborative knowledge advancement and socio-emotional interactions. The experimental group (n = 48), utilizing MASRA with visualized analytics and reflective prompts, was compared to a comparison class (n = 38) using portfolio assessment. Analysis of 4034 discussion notes revealed the MASRA group significantly outperformed the comparison class across six collaboration dimensions, achieving nearly double the collective knowledge advancement. The experimental class demonstrated more positive and fewer negative socio-emotional interactions, alongside more frequent "negative → positive" transitions. Furthermore, MASRA strengthened the co-occurrence of metacognitive discourse and positive socio-emotional dynamics, suggesting it enhances students' ability to regulate emotions and maintain group cohesion. These findings indicate that MASRA is a scalable model that effectively fosters both cognitive-level knowledge innovation and productive socio-emotional dynamics in collaborative learning environments, offering evidence-based strategies for higher education.

Knowledge building (KB) is a deep constructivist teaching model emphasizing collective responsibility for idea improvement and community development. Technology-supported KB has shown promise in fostering higher-order skills among university students, including critical thinking, metacognition, and disciplinary literacy (Bereiter & Scardamalia, 2003; Moss & Beatty, 2006; Scardamalia, 2006; Yang, van Aalst, et al., 2020). However, realizing this potential depends heavily on students' social interactions, particularly socio-emotional interactions, which involve behaviors such as expressing care, developing trust, fostering belonging, and promoting a sense of community (Bakhtiar et al., 2018; Kreijns et al., 2003). Effective socio-emotional interactions are critical in collaborative inquiry, as they support team functioning, motivation, and cognitive engagement (Näykki et al., 2014; Pekrun & Perry, 2013; Pekrun, Vogl, et al., 2017). Positive socio-emotional interactions facilitate conflict resolution, help-seeking, collective planning, and behavioral engagement, whereas negative interactions—arising from cognitive challenges, motivational issues, or interpersonal dynamics—can impede learning and reduce satisfaction (Blumenfeld et al., 1996; Kempler & Linnenbrink, 2013; Lajoie et al., 2015; Näykki et al., 2014). Despite this, research on promoting socio-emotional interactions using technology remains limited, especially in complex, open-ended learning contexts.

Reflective assessment involves monitoring, reflecting on, and adjusting one's learning using evidence from rubrics or analytics, fostering metacognitive awareness and self-regulation (Yang, Zhu, et al., 2022). Multimodal analytics-supported reflective assessment (MASRA) extends this approach by providing learners with real-time, multimodal data—behavioral and self-report—on their cognitive, metacognitive, and emotional processes (Blikstein & Worsley, 2016; Tao & Zhang, 2021; Yang, Chen, et al., 2020; Yang, Zhu, et al., 2022;). MASRA integrates planning, monitoring, reflecting, and regulating, enabling students to identify challenges, adjust strategies, and transform negative socio-emotional experiences into constructive learning signals, while reinforcing positive affect and self-efficacy (Panadero, 2017; Efklides, 2011). Previous studies confirm MASRA's effectiveness in enhancing cognitive engagement and agency, particularly for low-achieving students in knowledge-building environments (Yang, 2019). However, its role in fostering socio-emotional interactions and the underlying mechanisms remain underexplored.

Socio-emotional interactions play a crucial role in collaborative learning. Positive interactions—such as mutual respect, support, and care—are associated with deeper cognitive engagement, successful problem-solving, and higher motivation (Rogat & Linnenbrink-Garcia, 2011; Näykki et al., 2014; Lajoie et al., 2015; Mullins, Deiglmayr, & Spada, 2013). They also facilitate monitoring information exchange, inclusive planning, and team behavioral engagement (Rogat & Adams-Wiggins, 2014; Kempler & Linnenbrink, 2013). Negative interactions, however, can stem from cognitive, motivational, or interpersonal challenges and may reduce help-giving, impede learning progress, and generate persistent negative emotions (Andriessen et al., 2013; Järvenoja & Järvelä, 2009;

 © ISLS

Näykki et al., 2014). Recent research has highlighted the importance of multimodal analytics in capturing these complex social-emotional dynamics, offering richer insights than single-source measures (Isohätälä et al., 2020; Huang & Lajoie, 2023).

Socio-emotional dynamics are often subtle and difficult for learners to perceive during ongoing inquiry. By integrating multiple data streams, MASRA externalizes these tacit interaction patterns, reducing the cognitive burden of self-diagnosing group climate. Visualizing these dynamics provides concrete evidence that prompts learners to enact repair-oriented transitions, linking metacognitive reflection directly to collaborative regulation. This aligns with recent CSCL insights on how carefully designed analytics-supported dashboards foster reflective assessment (Gutiérrez-Braojos et al., 2025). This approach not only supports individual learning but also cultivates a shared sense of community, collaborative planning, and collective responsibility (Scardamalia & Bereiter, 2014).

The present study addresses a gap in the literature by exploring how MASRA can support productive socio-emotional interactions in university KB contexts. We designed teaching activities and reflective materials to generate rich multimodal interaction data, which were analyzed using content analysis, lag sequential analysis, and cognitive network visualization. The study investigates (1) the effects of MASRA on socio-emotional interactions among undergraduates, (2) differences in socio-emotional dynamics across students with varying academic performance, and (3) the relationship between socio-emotional interactions and individual learning outcomes in collaborative knowledge construction. By integrating reflective assessment with multimodal analytics, this research provides empirical insights into fostering positive social-emotional engagement and enhancing both cognitive and emotional dimensions of learning in higher education. The following research questions were investigated:

RQ 1. Do the undergraduates with multimodal analytics-supported reflective assessment design perform better in advancing collective knowledge?

RQ 2. Do the undergraduates with multimodal analytics-supported reflective assessment design engaged more in productive socio-emotion and socio-emotional interaction?

RQ 3. How do the undergraduates promote positive socio-emotional interactions through the enactment of multimodal analytics-supported reflective assessment?

Methods

Research context and participants

The study involved 86 second-year undergraduate students majoring in Educational Technology at a normal university in central China. They participated in a compulsory 18-week course (two 45-minute lessons per week) designed to engage students in expert-like problem discovery and problem-solving within a technology-supported knowledge-building (KB) environment. The course aimed to develop higher-order skills such as collaboration, inquiry, metacognition, and knowledge building, while promoting productive socio-emotional interactions. A quasi-experimental design was adopted: the experimental class (n = 48) experienced KB augmented by MASRA, whereas the comparison class (n = 38) engaged in portfolio-supported reflective assessment. In the comparison condition, students manually selected representative Knowledge Forum notes, justified their choices, analyzed inquiry strengths, and synthesized themes without data-driven visualizations. This contrast allowed us to isolate the effects of offloading the cognitive demands of reflection onto structured multimodal analytics. The study received ethical approval from the university's review board prior to data collection. All participants provided informed consent and could withdraw at any time.

KB design augmented by multimodal analytics-supported reflective assessment

The instructional design (Figure 1) integrated a data-driven three-element reflective assessment model, inspired by van Aalst and Chan's four-phase KB framework and Yang et al.'s three-phase KB model (van Aalst & Chan, 2011; Yang, Chen, et al., 2020). In Phase 1, students established supportive social relationships and a sense of teamwork and responsibility. Phase 2 (Weeks 3–8) involved collaborative KB activities focused on creative knowledge through face-to-face and online discussions and reflections. In Phase 3 (Weeks 9-18), MASRA provided three specific analytics outputs via dashboards shown in Figure 3: (1) radar charts of collaborative KB behaviors generated via GCA, (2) KBDeX social network visualizations of idea connections, and (3) sequential transition graphs of socio-emotional patterns. To ensure replicability, the instructor first modeled data interpretation, explicitly framing analytics for regulation rather than evaluation. Groups then followed a fixed sequence—diagnosing KB performance prior to examining socio-emotional conditions—guided by structured metacognitive prompt sheets organized as a reflection cycle.

Data sources and analysis

To examine the effects of multimodal analytics-supported reflective assessment (MASRA) on undergraduates' socio-emotional interactions and collective knowledge building (see Figure 2), we collected 2,603 Knowledge Forum notes from the experimental class and 1,431 notes from the comparison class. Notes were preprocessed and analyzed using content analysis, lag sequential analysis (LSA), and epistemic network analysis (ENA).

The impacts of multimodal analytics-supported reflective assessment on undergraduates' improvement of collective knowledge (RQ1)

Group Communication Analysis (GCA) was employed to assess collaborative knowledge building. GCA combines computational linguistic techniques with sequential interaction analysis to capture micro-level patterns of group communication. Six dimensions—participation, internal cohesion, overall responsivity, social impact, newness, and communication density—were used as indicators of collective knowledge advancement. Independent sample t-tests compared the experimental and comparison classes to examine the effect of MASRA on knowledge construction.

Figure 1

Instructional Design for Knowledge Construction with Three Elements of Data-driven Reflective Evaluation

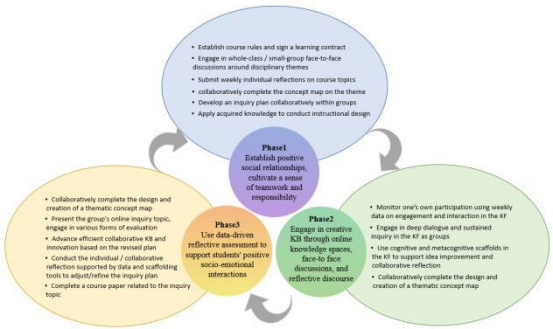


Figure 2

The Analytical Methods

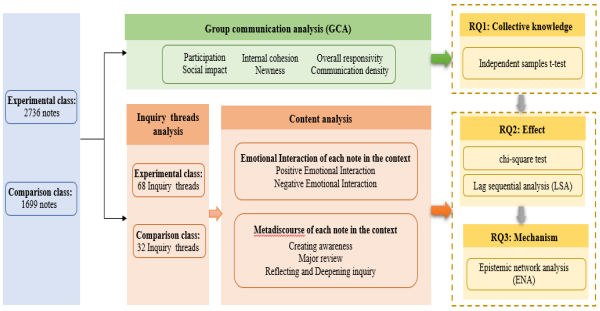
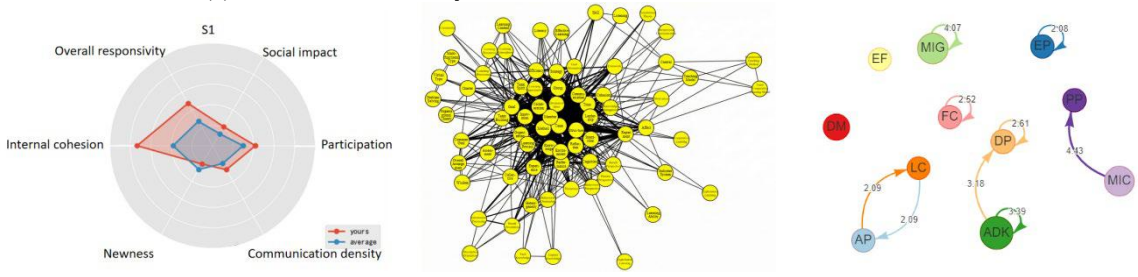


Figure 3

MASRA Dashboard Examples Provided to Students: (a) Collaborative KB Behaviors Radar Chart, (b) KBDeX Idea Network, and (c) Socio-emotional Sequential Transition Pattern.



(a) (b) (c)

Effect of multimodal analytics-supported reflective assessment on the undergraduates' productive socio-emotional interactions (RQ2)

Students' notes were first organized into distinct inquiry threads, with each thread containing a cluster of notes addressing the same inquiry problem. Notes were grouped according to their logical sequence and timestamps. In total, 32 inquiry threads (1,431 notes) were identified in the comparison class, whereas 68 threads (2,603 notes) appeared in the experimental class.

A content analysis framework was developed to examine students' socio-emotional interactions. The framework drew on the work of Kempler et al. (2013) and Rogat and Adams-Wiggins (2014) to identify basic types of socio-emotional interaction. Subdimensions of positive socio-emotional interaction were further refined using the concepts of emotional, interactive, and cohesive social presence proposed by Lajoie et al. (2015). Each note was coded into a single socio-emotional category, with the dominant type selected when multiple expressions occurred. Two researchers independently coded over 20% of the notes (experimental: n = 550; comparison: n =

340), achieving high interrater reliability (Cohen's κ = 0.891 for experimental, 0.853 for comparison). Discrepancies were resolved through discussion, and one researcher coded the remaining notes.

To examine changes in socio-emotional interactions and the effects of multimodal analytics-supported reflective assessment, chi-square tests were conducted to compare interaction patterns between classes and across pre-intervention and mid-intervention phases. Lag sequential analysis (LSA) was employed to investigate transition patterns of socio-emotional interactions. Using GSEQ software, Z-scores were calculated for each transition, with values above 1.96 indicating statistically significant sequences (Bakeman & Quera, 2011).

Comparison Between Classes and Over Time

LSA compared socio-emotional interaction sequences between the experimental and comparison classes. In addition, notes from the experimental class were divided into two phases—before and during MASRA implementation—to examine temporal changes. This analysis revealed how MASRA influenced the development of productive socio-emotional interactions among undergraduates.

Mechanism of multimodal analytics-supported reflective assessment for promoting undergraduates' productive socio-emotional interactions (RQ3)

ENA was used to investigate co-occurrence patterns between metacognitive utterances and socio-emotional interactions, visualizing and quantifying relationships that reveal MASRA's mechanism in promoting productive interactions. Epistemic networks were generated using the ENA web tool, and co-occurrence coefficients were calculated in RStudio to examine changes in social network characteristics across phases in the experimental class. This multi-method approach allowed us to capture both the quantitative improvements in collaborative knowledge and the qualitative shifts in socio-emotional interactions, providing insight into how MASRA fosters productive emotional engagement and supports higher-order collaborative processes.

Results

The impacts of multimodal analytics-supported reflective assessment on undergraduates' improvement of collective knowledge (RQ1)

Group Communication Analysis (GCA) framework developed by Dowell et al. (2019) evaluates learners' knowledge-building behaviors through semantic analysis of their contributions. It includes six key indicators: Participation, Internal Cohesion, Overall Responsivity, Social Impact, Newness, and Communication Density.These indicators reflect collective knowledge advancement.

As shown in Table 1, the experimental class outperformed the comparison class across all six dimensions, with differences reaching high statistical significance ($p < 0.01$). In particular, Internal Cohesion, Overall Responsivity, and Social Impact showed the largest effects. On several dimensions, the experimental class achieved average scores roughly twice as high as the comparison class. These results indicate that the integration of multimodal analytics-supported reflective assessment effectively enhanced students' collaborative knowledge-building behaviors and strengthened the collective advancement of ideas.

Table 1

GCA Results of Experimental Class and Comparison Class

	Class ($M \pm SD$)		t	p
	Experimental class (n = 48)	Comparison class (n = 38)		
Participation	52.98 ± 18.33	36.53 ± 12.31	-4.963	0.000**
Internal Cohesion	1254.61 ± 876.66	573.31 ± 364.65	-4.878	0.000**
Overall Responsivity	1167.63 ± 595.38	491.10 ± 173.74	-7.480	0.000**
Social Impact	1167.63 ± 615.05	492.44 ± 167.10	-7.274	0.000**
Newness	60.61 ± 23.68	42.02 ± 21.05	-3.795	0.000**
Communication Density	1377.85 ± 968.10	708.09 ± 458.43	-4.231	0.000**

* $p<0.05$ ** $p<0.01$

The effects of multimodal analytics-supported reflective assessment on undergraduates' productive socio-emotional interactions (RQ2)

Table 2 presents the overall distribution of socio-emotional interactions across the two classes. A chi-square test revealed significant differences between the MASRA and comparison conditions ($\chi^2 = 157.822$, $p < .001$). The MASRA class demonstrated higher proportions of productive moves, such as encouraging participation (19.53%

 © ISLS

vs. 15.94%) and fostering cohesion (35.26% vs. 25.24%), alongside significantly fewer negative patterns like passive participation (1.77% vs. 7.20%).

Sequential transitions of socio-emotional interactions in the experimental and comparison classes. We used the Viswork package for R to visualize significant transitions (i.e., a z-score greater than 1.96) in the two classes, as shown in Fig. 4. The greater the z-score, the stronger the link between the two nodes. The size of each socio-emotional interactions node denotes the frequency of the socio-emotional interactions, with a larger node representing a higher frequency.

As Figure 4 shows, there were significant differences between the experimental and comparison classes in terms of socio-emotional interaction transitions. The experimental class exhibited more transitions from negative to positive emotional interactions than comparison class, such as passive participation -> active participation (z-score = 3.1), low cohesion->mutual improvement through guidance (z-score = 2.25), and passive participation->demarginalization of members (z-score = 2.55). Meanwhile, the number of significant sequential transitions between negative socio-emotional interactions in the experimental class was markedly reduced compared to the comparison class, with six transition patterns disappearing entirely—particularly four transitions between low cohesion and misattribution of incompetence (low cohesion -> low cohesion, z-scores=6.06; low cohesion -> misattribution of incompetence, z-scores=4.08; misattribution of incompetence -> low cohesion, z-scores=4.08; misattribution of incompetence -> misattribution of incompetence, z-scores=4.18). These findings indicate that multimodal analytics-supported reflective assessment can enhance students' engagement in positive emotional interactions within the experimental class.

Table 2

Comparison of Socio-Emotional Interaction Behaviors between the Two Classes

	Experimental class		Comparison class	
	f	%	*f*	%
Positive Emotional Interaction				
active participation	681	26.18	428	29.93
encouraging participation	508	19.53	228	15.94
mutual improvement through guidance	79	3.04	20	1.40
authoritative discipline knowledge	265	10.19	194	13.57
fostering cohesion	917	35.26	361	25.24
demarginalization of members	45	1.73	32	2.24
Negative Emotional Interaction				
discouraging participation	8	0.31	10	0.70
low cohesion	4	0.15	11	0.77
misattribution of incompetence	16	0.62	24	1.68
passive participation	46	1.77	103	7.20
external factors	32	1.23	19	1.33

Undergraduates' engagement in productive socio-cognitive and socio-emotional interactions through multimodal analytics-supported reflective assessment (RQ3)

To answer RQ3, we compared the epistemic network characteristics of metacognitive utterances and socio-emotional interactions between the two classes, as well as across the two phases within the experimental class. This analysis aimed to explore the mechanism through which data-driven reflective assessment supports productive socio-emotional interactions.

Characterization of the socio-cognitive network of the experimental and comparison classes. Figure 5 displays the plotted points and ENA means of the experimental class and the comparison class, the respective mean networks in the experimental and comparison classes, and the subtracted epistemic network of the two classes. The red and blue dots represented the centroids of each PST in the experimental class and the comparison class, respectively. The mean centroids of all dots in each class are shown as squares with a 95% confidence interval for each dimension represented by the rectangular outline.

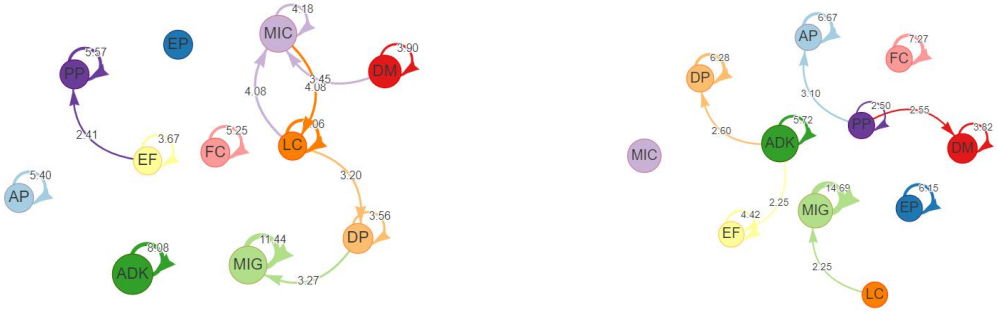
As shown in Figure 5(a) and Figure 5(c), the average centroid positions of students from the experimental and comparison classes differ in the cognitive network space, indicating a distinction between the two groups. Further independent sample t-tests revealed significant differences: in the dimension of positive socio-emotional interactions, the comparison class (M = -0.08, SD = 0.24) and the experimental class (M = 0.07, SD = 0.20) differed significantly along the Y-axis, $t(71.54) = 3.05$, $p < .001$; for negative emotional interactions, a significant difference was also observed along the X-axis between the comparison class (M = 0.16, SD = 0.26) and the experimental class (M = -0.13, SD = 0.32), $t(83.80) = -4.57$, $p < .001$.

 © ISLS

Moreover, as shown in Figure 5(b), Figure 5(d) the co-occurrence between metacognitive discourse and positive socio-emotional interactions was stronger in the experimental class than in the comparison class, for example, the connection between fostering cohesion and reflective data interpretation. In contrast, the cognitive network of the comparison class demonstrated stronger co-occurrence between metacognitive discourse and negative emotional interactions, such as the connection between passive participation and creating awareness, and the connection between low cohesion and monitoring and regulation. The results indicate that data-driven reflective assessment can enhance undergraduates' metacognitive discourse, which in turn promotes the co-occurrence of metacognitive discourse with positive socio-emotional interactions, thereby fostering the emergence of positive socio-emotional engagement among students.

Figure 4

Significant Sequential Transitions of Socio-emotional Interaction in the Two Classes.



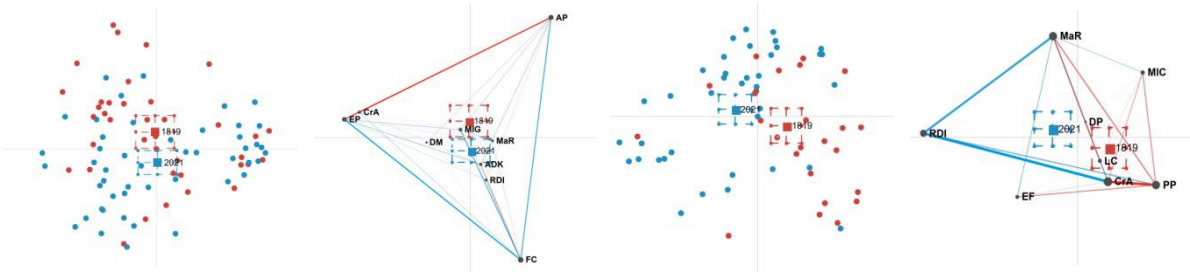
(a) (b)

Control class (a); Experimental class (b).

Changes in productive socio-cognitive and socio-emotional interactions in the experimental class over time. Figures 6(a) and 6(c) indicate that there were significant differences in the mean centroid values between the two phases of the experimental class. T-test results (assuming unequal variances) showed a significant difference in the Y-dimension for positive socio-emotional interactions between Phase 2 (M = 0.12, SD = 0.18) and Phase 1 (M = -0.12, SD = 0.28), $t(81.29) = 5.12$, $p < .001$; and a significant difference in the X-dimension for negative emotional interactions between Phase 2 (M = 0.16, SD = 0.26) and Phase 1 (M = -0.16, SD = 0.23), $t(93.14) = 6.38$, $p < .001$.

Figure 5

Means of Epistemic Network Analysis and Subtracted Network for the Experimental class and Comparison class.



(a) (b) (c) (d)

Comparisons of positive sentiment in KB actions for the experimental class (Blue) and comparison class (Red) using plotted points and centroids (a); Subtracted epistemic network of positive emotional interactions: the experimental class (Blue) and comparison class (Red) (b); Comparisons of negative sentiment in KB actions for the experimental class (Blue) and comparison class (Red) using plotted points and centroids (c); subtracted epistemic network of negative emotional interactions: the experimental class (Blue) and comparison class (Red) (d).

As shown in Figures 6(b), 6(d) at the level of positive socio-emotional interaction, in Phase 2, students in the experimental class established more connections at the misattribution of incompetence and fostering

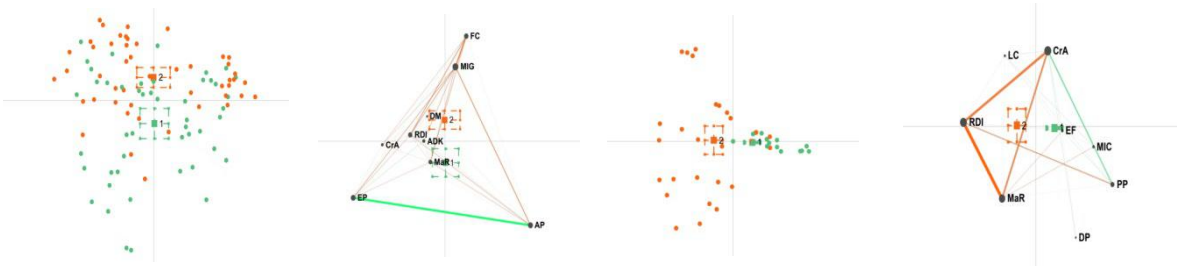
cohesion nodes. Additionally, compared to Phase 1, stronger co-occurrence was observed between metacognitive discourse and positive socio-emotional interactions, for example, the connection between fostering cohesion and monitoring and regulation and the connection between misattribution of incompetence and reflective data interpretation. In contrast, at the level of negative socio-emotional interaction, Phase 1 showed higher co-occurrence strength among negative emotional interactions, such as, the connection between passive participation and external factors.

Changes in productive socio-cognitive and socio-emotional interactions in the comparison class over time. Figures 7(a) and 7(c) reveal significant differences in the mean centroid values across the two phases in the comparison class. T-test results (assuming unequal variances) showed a significant difference in the X-dimension ($t(63.79) = 3.01$, $p < .001$) and Y-dimension ($t(70.42) = 2.62$, $p < .001$) for positive socio-emotional interactions between Phase 2 and Phase 1; and a significant difference in the X-dimension for negative emotional interactions between Phase 2 ($M = -0.02$, $SD = 0.24$) and Phase 1 ($M = -0.12$, $SD = 0.28$), $t(71.41) = 3.29$, $p < .001$.

In the second phase, the co-occurrence coefficients between positive socio-emotional interactions and metacognition among undergraduates were noticeably higher than in the first phase. Notably, MaR and RDI co-occurred with various positive socio-emotional interactions only in the second phase. At the level of negative socio-emotional interactions, the overall co-occurrence coefficients between such interactions showed a declining trend. For example, LC–MIC (0.03 vs. 0.00) and PP–EF (0.08 vs. 0.03) both exhibited decreases.

Figure 6

Means of Epistemic Network Analysis and Subtracted Network for the Experimental Class Phase 1 and Phase 2.

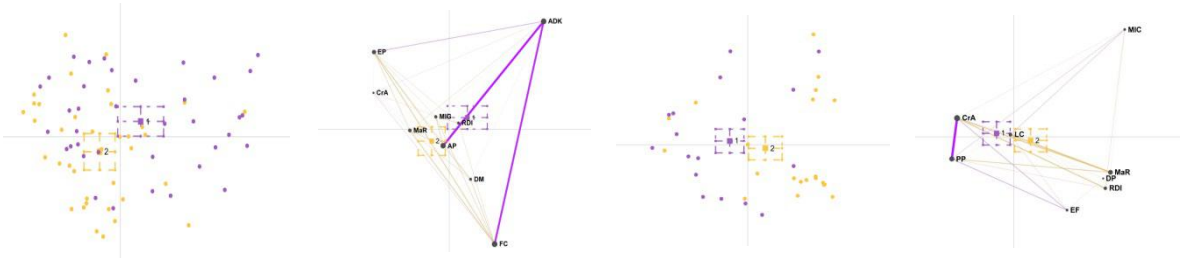


(a) (b) (c) (d)

Comparisons of positive sentiment in KB actions for the phase 1 (Green) and phase 2(Orange) using plotted points and centroids (a); Subtracted epistemic network of positive emotional interactions: the phase 1 (Green) and phase 2(Orange) (b); Comparisons of negative sentiment in KB actions for the phase 1 (Green) and phase 2(Orange) using plotted points and centroids (c) Subtracted epistemic network of negative emotional interactions: the phase 1 (Green) and phase 2(Orange) (d).

Figure 7

Means of Epistemic Network Analysis and Subtracted Network for the Control Class Phase 1 and Phase 2.



(a) (b) (c) (d)

Comparisons of positive sentiment in KB actions for the phase 1 (Purple) and phase 2(Yellow) using plotted points and centroids (a); Subtracted epistemic network of positive emotional interactions: the phase 1 (Purple) and phase 2(Yellow) (b); Comparisons of negative sentiment in KB actions for the phase 1 (Purple) and phase 2(Yellow) using plotted points and centroids (c); Subtracted epistemic network of negative emotional interactions: the phase 1 (Purple) and phase 2(Yellow)(d).

Discussion and conclusions

Impacts on collective knowledge (RQ1)

Using Group Communication Analysis (GCA), this study examined how multimodal analytics-supported reflective assessment influenced undergraduates' collaborative knowledge building. GCA evaluates collaboration across six dimensions—Participation, Internal Cohesion, Overall Responsivity, Social Impact, Newness, and Communication Density—capturing key cognitive behaviors. The results showed significant improvements across all dimensions in the experimental class compared with the control class, confirming that reflective assessment can enhance participation, collective epistemic agency, and the advancement of shared knowledge (Yang, Yuan, et al., 2022; Yang, Chen, et al., 2024; Yang, Feng, et al., 2024).

Impacts on productive socio-emotional interactions (RQ2)

Data-driven reflective assessment also promoted positive socio-emotional interactions. Students in the experimental class exhibited higher proportions of supportive and cohesive behaviors and demonstrated more positive-to-positive and negative-to-positive interaction sequences over time. These findings extend previous work showing that reflective assessment enhances cognitive engagement and epistemic emotions (Yang, Yuan, et al., 2024), revealing that it can also scaffold beneficial socio-emotional dynamics. By integrating visualized dashboards and reflective prompt sheets, students were able to monitor their knowledge and regulate emotions, supporting proactive collaboration and a shared sense of community.

Mechanisms of data-driven reflective assessment (RQ3)

Epistemic network analysis revealed stronger co-occurrence patterns between metacognitive discourse and positive socio-emotional interactions in the experimental group, particularly during the second phase of the intervention. These results suggest that reflective assessment enables students to actively monitor and reflect on their socio-emotional states, plan strategies to address negative interactions, and maintain productive group dynamics. Consequently, negative interactions can be transformed into positive ones, highlighting the interplay among metacognition, socio-emotional engagement, and collaborative knowledge construction.

Practical implications

This study offers several implications for designing technology-supported learning environments. First, embedding data-driven reflective assessment within a collaborative culture fosters positive socio-emotional interactions and collective responsibility. Second, reflective prompts and visualized dashboards help students interpret learning analytics, monitor progress, and regulate both cognitive and emotional processes. Third, structured activities can guide students in identifying and addressing negative interactions, supporting constructive socio-emotional dynamics. Overall, positive socio-emotional interactions are crucial not only for learning outcomes but also for sustained engagement and higher-order skill development.

Limitations and future directions

The study has several limitations. First, the lack of a pre-test limits the precision of measuring knowledge gains. Future studies should include pre- and post-tests to assess learning progress. Second, while multiple data sources were collected, this study focused mainly on knowledge forum notes. Future research should leverage other artifacts (e.g., reflective sheets, audio/video recordings) and employ qualitative analyses to explore mechanisms more deeply. Third, manual coding of discourse, while rich in context, is time-consuming and limits scalability. Automated methods for analyzing socio-emotional interactions could enhance efficiency and support real-time feedback.

Conclusions

This study demonstrates that multimodal analytics-supported reflective assessment fosters productive socio-emotional interactions and advances collective knowledge in a knowledge-building environment. By integrating scaffolds and reflective prompts, students monitored their socio-emotional states, engaged in metacognitive regulation, and enhanced collaboration. The findings extend KB theory by showing that reflective assessment influences both cognitive and socio-emotional processes. Practically, this study offers a replicable three-phase instructional model and evidence-based strategies for using analytics to cultivate socio-emotional skills, metacognitive regulation, and collaborative knowledge creation. The results also clarify the mechanism: data-driven reflection allows students to transform negative interactions into positive ones, revealing the interconnected roles of instructional practices, socio-emotional dynamics, and reflective assessment in supporting productive collaborative learning.

 © ISLS

References

Andriessen, J., Pardijs, M., & Baker, M. J. (2013). Getting on and getting along: Tension in the development of collaboration. In M. J. Baker, J. Andriessen, & S. Järvelä (Eds.), *Affective learning together: Social and emotional dimensions of collaborative learning* (pp. 205–230). Routledge.

Bakeman, R., & Quera, V. (2011). *Sequential analysis and observational methods for the behavioral sciences*. Cambridge University Press.

Bakhtiar, A., Webster, E. A., & Hadwin, A. F. (2018). Regulation and socio-emotional interactions in a positive and a negative group climate. *Metacognition and Learning*, *13*(1), 57–90.

Bereiter, C., & Scardamalia, M. (2003). Learning to work creatively with knowledge. *Powerful learning environments: Unravelling basic components and dimensions* (pp. 55–68). Pergamon.

Blikstein, P., & Worsley, M. (2016). Multimodal learning analytics and education data mining: Using computational technologies to measure complex learning tasks. *Journal of Learning Analytics*, *3*(2), 220–238.

Blumenfeld, P., Marx, R., Soloway, E., & Krajcik, J. (1996). Learning with peers: From small group cooperation to collaborative communities. *Educational Researcher*, *25*(8), 37–40.

Dowell, N. M., Nixon, T. M., & Graesser, A. C. (2019). Group communication analysis: A computational linguistics approach for detecting sociocognitive roles in multiparty interactions. *Behavior Research Methods*, *51*(3), 1007–1041.

Efklides, A. (2011). Interactions of metacognition with motivation and affect in self-regulated learning: The MASRL model. *Educational Psychologist*, *46*(1), 6–25.

Gutiérrez-Braojos, C., Rodríguez-Domínguez, C., de Luna, E. B., Martínez, M., Colas, M. P., & Hernández, F. M. (2025). Adaptive dashboard system for reflective assessment in knowledge building pedagogy. In *Proceedings of the 18th International Conference on Computer-Supported Collaborative Learning-CSCL 2025* (pp. 602–604). International Society of the Learning Sciences.

Huang, X., & Lajoie, S. P. (2023). Social emotional interaction in collaborative learning: Why it matters and how can we measure it? *Social Sciences & Humanities Open*, *7*(1), 100447.

Isohätälä, J., Näykki, P., & Järvelä, S. (2020). Cognitive and socio-emotional interaction in collaborative learning: Exploring fluctuations in students' participation. *Scandinavian Journal of Educational Research*, *64*(6), 831–851.

Järvenoja, H., & Järvelä, S. (2009). Emotion control in collaborative learning situations: Do students regulate emotions evoked by social challenges? *British Journal of Educational Psychology*, *79*, 463–481.

Kempler, T. M., & Linnenbrink, E. A. (2013). Helping behaviors in collaborative groups in math: A descriptive analysis. In *Help seeking in academic settings* (pp. 89–115). Routledge.

Kreijns, K., Kirschner, P. A., & Jochems, W. (2003). Identifying the pitfalls for social interaction in computer-supported collaborative learning environments: A review of the research. *Computers in Human Behavior*, *19*(3), 335–353.

Lajoie, S. P., Lee, L., Poitras, E., Bassiri, M., Kazemitabar, M., Cruz-Panesso, I., ... Lu, J. (2015). The role of regulation in medical student learning in small groups: Regulating oneself and others' learning and emotions. *Computers in Human Behavior*, *52*, 601–616.

Moss, J., & Beatty, R. (2006). Knowledge building in mathematics: Supporting collaborative learning in pattern problems. *International Journal of Computer-Supported Collaborative Learning*, *1*, 441–465.

Mullins, D., Deiglmayr, A., & Spada, H. (2013). Motivation and emotion shaping knowledge co-construction. In *Affective learning together* (pp. 139–161). Routledge.

Näykki, P., Järvelä, S., Kirschner, P. A., & Järvenoja, H. (2014). Socio-emotional conflict in collaborative learning—A process-oriented case study in a higher education context. *International Journal of Educational Research*, *68*, 1–14.

Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology*, *8*, 422.

Pekrun, R., & Perry, R. P. (2013). Control-value theory of achievement emotions. In *International handbook of emotions in education*. Routledge.

Pekrun, R., Vogl, E., Muis, K. R., & Sinatra, G. M. (2017). Measuring emotions during epistemic activities: The Epistemically-Related Emotion Scales. *Cognition and Emotion*, *31*(6), 1268–1276.

Rogat, T. K., & Adams-Wiggins, K. R. (2014). Other-regulation in collaborative groups: Implications for regulation quality. *Instructional Science*, *42*(6), 879–904.

Rogat, T. K., & Linnenbrink-Garcia, L. (2011). Socially shared regulation in collaborative groups: An analysis of the interplay between quality of social regulation and group processes. *Cognition and Instruction*, *29*(4), 375–415.

Scardamalia, M. (2006). Knowledge building: Theory, pedagogy, and technology. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 97–118). Cambridge University Press.

Scardamalia, M., & Bereiter, C. (2014). Knowledge building and knowledge creation: Theory, pedagogy, and technology. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (2nd ed., pp. 397–417). Cambridge University Press.

Tao, D., & Zhang, J. (2021). Agency to transform: How did a grade 5 community co-configure dynamic knowledge building practices in a yearlong science inquiry? *International Journal of Computer-Supported Collaborative Learning*, *16*(3), 403–434.

van Aalst, J., & Chan, C. K. (2011). Empowering students as knowledge builders. In *Transformative approaches to new technologies and student diversity in futures oriented classrooms: Future proofing education* (pp. 85–103). Springer.

Yang, Y. (2019). Reflective assessment for epistemic agency of academically low-achieving students. *Journal of Computer Assisted Learning*, *35*(4), 459–475.

Yang, Y., Chen, Q., Yu, Y., Feng, X., & van Aalst, J. (2020). Collective reflective assessment for shared epistemic agency by undergraduates in knowledge building. *British Journal of Educational Technology*, *51*(4), 1136–1154.

Yang, Y., Chen, Y., Feng, X., Sun, D., & Pang, S. (2024). Investigating the mechanisms of analytics-supported reflective assessment for fostering collective knowledge. *Journal of Computing in Higher Education*, *36*(1), 242–273.

Yang, Y., Feng, X., Zhu, G., & Xie, K. (2024). Effects and mechanisms of analytics-assisted reflective assessment in fostering undergraduates' collective epistemic agency in computer-supported collaborative inquiry. *Journal of Computer Assisted Learning*, *40*(3), 1098–1122.

Yang, Y., van Aalst, J., & Chan, C. K. (2020). Dynamics of reflective assessment and knowledge building for academically low-achieving students. *American Educational Research Journal*, *57*(3), 1241–1289.

Yang, Y., Yuan, K., Feng, X., Li, X., & van Aalst, J. (2022). Fostering low-achieving students' productive disciplinary engagement through knowledge-building inquiry and reflective assessment. *British Journal of Educational Technology*, *53*(6), 1511–1529.

Yang, Y., Yuan, K., Zhu, G., & Jiao, L. (2024). Collaborative analytics-enhanced reflective assessment to foster conducive epistemic emotions in knowledge building. *Computers & Education*, *209*, 104950.

Yang, Y., Zhu, G., Sun, D., & Chan, C. K. (2022). Collaborative analytics-supported reflective assessment for scaffolding pre-service teachers' collaborative inquiry and knowledge building. *International Journal of Computer-Supported Collaborative Learning*, *17*(2), 249–292.

 © ISLS

International Society of the Learning Sciences

Effects and Mechanisms of Learning Analytic-Supported Reflective Assessment for Promoting Metacognitive Regulation among Undergraduates

Yuqin Yang, Central China Normal University, China, yangyuqin@ccnu.edu.cn
Yangbo Zhao, Central China Normal University, China, zhaoyangbo@mails.ccnu.edu.cn
Xueqi Feng, Southern University of Science and Technology, China, fengxq@sustech.edu.cn
Daner Sun, The Education University of Hong Kong, Hong Kong S.A.R., dsun@eduhk.hk
Jianwen Sun, Central China Normal University, China, sunjw@ccnu.edu.cn

Abstract: Metacognitive regulation is critical for effective learning, yet many undergraduates lack the necessary skills to self-regulate. This study conducted an 18-week quasi-experimental investigation with 88 undergraduates in a Knowledge Building (KB) environment to examine how learning analytics-supported reflective assessment (LAsRA) influences collaborative knowledge creation and metacognitive regulation. Group Communication Analysis (GCA) and Epistemic Network Analysis (ENA) were employed to analyze students' discourse and regulatory networks. GCA results showed that the experimental group significantly outperformed the control group in knowledge creation, exhibiting highly cohesive discourse. ENA revealed that LAsRA facilitated a critical shift toward socially shared and deep-level metacognitive regulation. Rather than merely monitoring outcomes, students successfully constructed a dynamic collaborative cycle of "goal-setting—process monitoring—reflective evaluation — dynamic regulation." Findings suggest that analytics-supported reflective assessment provides an effective pathway to foster a community-level, self-improving regulatory environment, fundamentally enhancing students' metacognitive capabilities and collective knowledge advancement.

Introduction

In today's knowledge-driven society, higher education increasingly aims to cultivate students' higher-order competencies to meet future challenges, including metacognitive regulation, inquiry, collaboration, and knowledge creation (Scardamalia & Bereiter, 2014; Snell & Lefstein, 2018). In collaborative learning contexts, metacognitive regulation is considered the core of effective learning, requiring learners to strategically coordinate cognitive, motivational, emotional, and behavioral processes to achieve task goals (De Backer et al., 2015; Järvelä et al., 2016). To cultivate these higher-order competencies, Knowledge Building (KB) offers a student-centered framework for collaborative inquiry (Scardamalia & Bereiter, 2006). In KB environments, students are not isolated individuals but active knowledge constructors who must assume collective cognitive responsibility for advancing the frontiers of community knowledge (Scardamalia, 2002). This advancement of knowledge requires students' deep, active participation across cognitive, metacognitive, social, and emotional dimensions, thereby extending the regulatory process from the individual's internal psychological level to the level of interpersonal interaction (Hadwin et al., 2011; Volet et al., 2009). Based on the level of social interaction, regulatory activities in collaborative contexts are conceptualized into three dynamically intertwined types: self-regulation (SR), co-regulation (CoR), and socially shared regulation (SSR) (Hadwin et al., 2011; Järvelä & Hadwin, 2013). Among these, SSR is the most critical for collaborative learning, extending regulation to the group level; it requires group members to interdependently negotiate shared task perceptions, establish collective goals, jointly monitor progress, and coordinate strategic adaptations to optimize collaborative performance (De Backer et al., 2015; Panadero & Järvelä, 2015).

Empirical studies suggest that high-level socially shared regulation is crucial for productive group engagement and sustained collaborative knowledge construction (Järvelä et al., 2016; Zheng et al., 2019). However, many undergraduates lack sufficient metacognitive awareness and skills to spontaneously initiate and maintain high-level SSR without the support of appropriate external scaffolding (De Backer et al., 2022; Raes et al., 2016). Collaborative inquiry is highly dynamic and complex; students must not only process domain knowledge content but also jointly monitor collective understanding and progress, which imposes a high cognitive load on learners (Raes et al., 2016; Zheng et al., 2019). When facing cognitive conflicts or challenges, the regulatory behaviors of groups lacking effective regulation often remain at the individual SR or CoR levels (Lee et al., 2015), easily triggering epistemic emotions such as confusion and anxiety (D'Mello et al., 2014; Pekrun et al., 2017; Yang, Yuan, et al., 2024). To address these challenges, prior research has introduced external scaffolds,

 © ISLS

such as collaboration scripts and metacognitive prompts, to guide students' regulatory processes, demonstrating positive effects on collaborative learning (Kim & Lim, 2018; Raes et al., 2016). Nevertheless, students often lack an intuitive perception of the dynamic evolution of collaboration, making it difficult to accurately identify the specific actions needed to resolve disagreements and advance collective knowledge. In such cases, their collaborative processes are prone to devolve into fragmented or short discussions lacking deep knowledge synthesis and substantive conceptual progress, forcing them to rely on low-level exploratory strategies (Yang et al., 2016; Zhang et al., 2018).

Reflective assessment has been proposed as an effective scaffolding strategy to support iterative metacognitive behaviors (White & Frederiksen, 1998; Yang et al., 2016). Reflective assessment prompts students to exercise agency and responsibility, actively examining their cognitive and metacognitive processes through cyclical activities of goal-setting, monitoring, reflection, and regulation, thereby enhancing regulatory awareness and skills, and developing internal metacognitive structures (Lei & Chan, 2018; Yang, Chen, et al., 2020). In recent years, with the rapid development of learning analytics technologies, reflective assessment has been further enhanced and expanded (Siemens & Gašević, 2012; Tao & Zhang, 2021). As a commonly used learning analytics tool, KBDeX can visualize complex collaborative interactions and knowledge-building processes, helping students utilize visualized charts and data to conduct reflective assessments (Oshima et al., 2012; Yang, van Aalst, & Chan, 2020). On the one hand, visualized collaborative data and analytics help students objectively evaluate group contributions, identify inquiry gaps, and intentionally create cognitive discrepancies between current inquiry states and desired goals (Yang et al., 2022). This not only effectively reduces the group's cognitive load when monitoring interaction processes but also transforms the confusion or frustration accompanying inquiry into positive epistemic emotions, such as curiosity, which drive deeper investigation (D'Mello et al., 2014; Yang, Yuan, et al., 2024). On the other hand, by integrating structured metacognitive prompts, learning analytics-supported reflective assessment (LAsRA) can guide student groups to engage in collective meaning negotiation and formulate evidence-based regulatory plans based on the interpretation of analytical data (Järvelä et al., 2015; Miller & Hadwin, 2015). Prior research indicates that this approach can effectively promote students' inquiry and the advancement of collective knowledge by alleviating associated cognitive burdens; moreover, even after the removal of learning analytics data and accompanying reflective scaffolds, students can continue to advance collective knowledge, suggesting that reflective assessment fosters self-sustaining metacognitive capabilities (Yang, Feng, et al., 2024). Therefore, the deep integration of learning analytics and reflective assessment holds the promise of empowering learners in highly complex collaborative environments, stimulating and sustaining a dynamic, community-level cycle of metacognitive regulation, thereby providing an effective practical pathway for shifting from individual regulation to high-level SSR (Zhang, Chen, et al., 2021).

Building on this background, the present study conducts an 18-week quasi-experimental investigation to deeply explore the effects of analytics-supported reflective assessment on undergraduates' collaborative knowledge building and metacognitive regulation (Yang et al., 2022). Using a multi-method approach, this study utilizes Group Communication Analysis (GCA) to objectively quantify socio-cognitive processes and knowledge-building outcomes, and employs Epistemic Network Analysis (ENA) to capture the network characteristics and co-occurrence structures of metacognitive regulatory behaviors. This study aims to address the following three specific research questions: (1) Do undergraduates with analytics-supported reflective assessment perform better in advancing collaborative knowledge building? (2) To what extent does reflective assessment affect metacognitive regulation? (3) How does reflective assessment support the mechanism of metacognitive regulation?

Methods

Research context and participants

This study was conducted in a compulsory core course, Scientific Inquiry and Knowledge Creation, at a comprehensive research university in Central China. The course, spanning 18 weeks with two 45-minute sessions per week, aimed to develop undergraduates' foundational knowledge and skills in scientific inquiry, critical thinking, and metacognitive regulation. It adopted a blended learning environment combining face-to-face and online collaborative inquiry. A total of 88 third-year students from diverse academic backgrounds participated, most with limited prior experience in scientific inquiry and knowledge creation. A quasi-experimental design was employed, with an experimental class (n = 50) and a comparison class (n = 38). Students worked in small teams of 4 to 6 members to investigate pre-assigned inquiry topics, with 10 groups in the experimental class and 7 groups in the comparison class. All participants provided informed consent and were assured of confidentiality and voluntary participation.

KB design enhanced by analytics-assisted reflective assessment

The instructional design integrated Knowledge Building (KB) principles—epistemic agency, reflective and transformative assessment, collective responsibility for idea improvement, and rise-above—into a three-component model to support collaborative knowledge creation and metacognitive regulation in a blended learning environment.

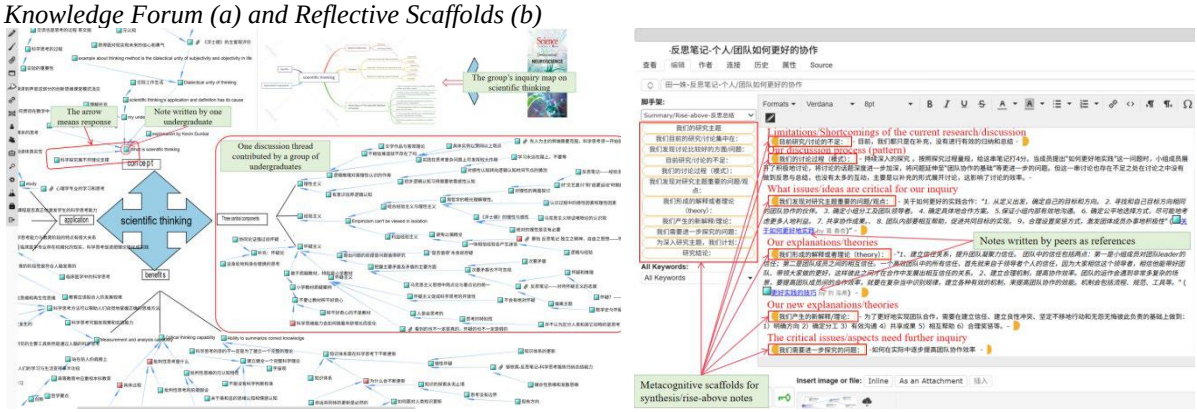
Component 1: Establishing a collaborative learning culture (Weeks 1–5). Students engaged in principle-based activities to cultivate awareness of collaborative inquiry and metacognition. Each student signed a learning contract that specified tasks, monitoring strategies, and weekly reflection topics. Guiding prompts (e.g., "What I have and haven't known about the topic, and what I want to know") supported reflective collaboration. Small groups investigated topics such as maintaining motivation and persistence in scientific inquiry, created posters, presented their findings, and developed research plans to guide their inquiry.

Component 2: Engaging in problem-centered inquiry (Weeks 4–8). Students conducted inquiries around selected core topics, posing questions, sharing ideas, and reflecting on discussions on Knowledge Forum (KF). Two types of scaffolds were provided in this phase: theory-building scaffolds (e.g., "My theory," "A better theory") to advance collective knowledge, and reflective scaffolds (e.g., "Our research topic," "Critical issues to explore") to support metacognitive development (see Figure 1). Students analyzed discussion notes and generated reflective summaries weekly using KF's visualization and integrated assessment tools.

Component 3: Deepening inquiry with analytics-supported reflective assessment (Weeks 8–18). As inquiry notes accumulated, the KBDeX tool was introduced in Week 9 to visualize the dynamic relationships of knowledge building. Students utilized the growth curve of cumulative degree centrality to monitor the progress of collective knowledge, and identified core concepts and promising areas for inquiry through the idea-network map. To support data interpretation, the instructor provided structured reflective prompt sheets (encompassing "our analysis," "our problems," and "our plan"), which guided students to identify cognitive gaps and plan subsequent inquiries. Meanwhile, they were encouraged to synthesize knowledge by creating rise-above notes and concept maps. In Week 13, a second round of reflective assessment was conducted, where students optimized their inquiry patterns by comparing their group's KBDeX data with that of the whole community. Furthermore, continuous weekly feedback on individual participation strengthened metacognitive monitoring and regulation throughout the entire process.

Figure 1

Knowledge Forum (a) and Reflective Scaffolds (b)



(a) (b)

Instruction in the comparison class

The comparison class, taught by the same instructor, engaged in identical activities during Components 1 and 2. The core instructional difference occurred in Component 3, where the comparison class advanced collective inquiry through two rounds of electronic portfolio-supported reflective assessment—an approach whose effectiveness has been confirmed by prior studies (Lei & Chan, 2018). In the first round, students reviewed their Knowledge Forum (KF) notes, selected high-rated clusters containing at least five notes, and analyzed them using knowledge building principles such as "idea improvement" and "rise above" to evaluate current discussions and formulate subsequent plans. In the second round, after updating their portfolios based on ongoing inquiry, students engaged in cross-group discussions to compare inquiry trajectories and integrated peer feedback to further refine their portfolios.

 © ISLS

Measures and analysis

To investigate the effects of learning analytics-supported reflective assessment (LAsRA) on undergraduates' collaborative knowledge construction and metacognitive regulation (RQ1, RQ2), and the mechanisms influencing metacognitive regulation (RQ3), we collected 1721 Knowledge Forum (KF) notes from the experimental class and 706 notes from the comparison class.

Effects on Collective Knowledge Building (RQ1). Students' online discussions were analyzed using Group Communication Analysis (GCA), an advanced computational linguistics method that models collaborative discourse as temporal-semantic networks(Dowell et al., 2019; Dowell & Poquet, 2021). GCA quantifies key socio-cognitive dimensions, including: (1) interpersonal interaction, measured by Overall Responsivity and Social Impact; (2) individual cognitive processes, indicated by Internal Cohesion; and (3) information-level contributions, captured by Participation, Newness, and Communication Density. This approach allows fine-grained, data-driven measurement of knowledge construction processes and outcomes. Independent samples t-tests compared the experimental and comparison classes on these dimensions.

Effects and Mechanisms on Metacognitive Regulation (RQ2 & RQ3). To examine metacognitive regulation, all KF notes were coded using a three-level framework adapted from prior research (Su et al., 2018; Zhang, Wen, & Liu, 2021). First, we identified the type of regulation: Self-Regulation, Co-Regulation, and Socially Shared Regulation. Second, we classified the process of regulation into Planning, Monitoring, Regulating, and Evaluating. Third, the focus of regulation was categorized into Task Understanding, Content Monitoring, Process Monitoring, Positive Emotion, Negative Emotion, Confusion, and Organization.

Two researchers independently coded 1,000 notes (41% of the data), achieving high inter-rater reliability (Cohen's $\kappa = 0.85$). We compared the proportions of metacognitive behaviors between classes and conducted Chi-square tests to examine significant differences.To explore mechanisms of metacognitive development, Epistemic Network Analysis (ENA) was applied using ENA Webkit (www.epistemicnetwork.org). ENA constructs dynamic networks visualizing co-occurrences of metacognitive elements, capturing both individual and group-level structures and strengths of connections.

Results

Effects of analytics-supported reflective assessment on collective knowledge building (RQ1)

GCA was used to examine the socio-cognitive processes of students' collaborative knowledge creation in the experimental (n = 50) and comparison (n = 38) classes. Independent samples t-tests revealed that the experimental group significantly outperformed the comparison group across all GCA dimensions ($p < .001$, see Table 1), indicating a comprehensive positive effect of analytics-supported reflective assessment on collaborative knowledge building.

Table 1

GCA Results of Experimental Class and Comparison Class

	Class ($M \pm SD$)			
	Experimental class (n = 50)	Comparison class (n = 38)	t	p
Participation	32.62 ±15.53	10.76 ± 7.27	-8.770	0.000**
Internal cohesion	499.41 ± 423.04	61.61 ± 100.56	-7.060	0.000**
Overall responsivity	413.15 ± 206.77	46.89 ± 38.84	-12.244	0.000**
Social impact	413.15 ± 248.04	46.89 ± 49.26	-10.181	0.000**
Newness	49.29 ± 34.51	22.49 ± 24.83	-4.054	0.000**
Communication density	548.86 ± 466.05	126.08 ± 95.70	-6.244	0.000**

* p<0.05 ** p<0.01

Specifically, experimental students demonstrated much higher participation, reflecting greater engagement and effort in discussions. Their Internal Cohesion and Overall Responsivity scores suggest more coherent individual contributions and stronger responsiveness to peers' ideas. Social Impact indicates that the experimental class exerted greater influence on collective discourse. Furthermore, higher Newness and Communication Density show that their contributions were both more innovative and efficiently articulated. These results suggest that integrating learning analytics with reflective assessment not only enhances students'

engagement and cognitive depth but also promotes more interactive, self-regulated, and socially regulated collaborative knowledge creation, leading to higher-quality contributions and richer group discourse compared with a standard Knowledge Building approach.

Effects of analytics-supported reflective assessment on metacognitive regulation (RQ2)

Proportions of metacognitive regulation in the experimental and comparison classes

Table 2 presents the proportions of metacognitive regulation behaviors for both classes, Chi-square tests revealed statistically significant differences across all regulation dimensions. Regarding regulation type ($\chi^2(df = 2, N = 2428) = 60.68$, $p < .001$), the experimental class exhibited a significantly higher proportion of SSR (55.43% vs. 38.61%), whereas the comparison class engaged more in CoR, this result indicates that experimental students' discussion was mainly about the social aspect of regulation. For the regulation process ($\chi^2(df = 3, N = 2428) = 21.18$, $p < .001$), the experimental class showed higher proportions of regulating (39.57% vs. 33.38%), evaluating (4.30% vs. 1.98%), and planning (8.72% vs. 8.06%). The comparison class primarily focused on monitoring the inquiry, exhibiting fewer effective evaluation behaviors. In terms of regulation focus ($\chi^2(df = 6, N = 2428) = 146.78$, $p < .001$), the experimental class demonstrated more process monitoring (24.40% vs. 7.21%) and organizing (4.30% vs. 2.40%), but less content monitoring (42.36% vs. 65.63%) than the comparison class. This indicates the experimental class consciously focused on the inquiry process itself, while the comparison class concentrated on knowledge content. Furthermore, the experimental class engaged more frequently in emotion regulation (e.g., positive emotion: 3.25% vs. 2.97%), which is noteworthy as managing confusion is a critical driver for deeper inquiry. These results suggest that learning analytics-supported reflective assessment prompted the experimental class to focus more on the regulation, evaluation, and emotional management of the KB process. To further explore the cyclical and interdependent characteristics of these behaviors, we subsequently employed ENA.

Table 2

Proportions of Metacognitive Regulation in the Experimental and Comparison Classes

	Experimental class		Comparison class		χ^2	*p*
	f	%	*f*	%		
Regulation type					60.68	0.00
Self-regulation	338	19.64	167	23.62		
Co-regulation	429	24.93	267	37.77		
Socially shared regulation	954	55.43	273	38.61		
Regulation process					21.18	0.00
Planning	150	8.72	57	8.06		
Monitoring	816	47.41	400	56.58		
Regulating	681	39.57	236	33.38		
Evaluating	74	4.30	14	1.98		
Regulation focus					146.78	0.00
Task understanding	400	23.24	150	21.22		
Content monitoring	729	42.36	464	65.63		
Process monitoring	420	24.40	51	7.21		
Positive emotion	56	3.25	21	2.97		
Negative emotion	7	0.41	1	0.14		
Confusion	35	2.03	3	0.42		
Organizing	74	4.30	17	2.40		

Epistemic networks of metacognitive regulation in the experimental and comparison classes

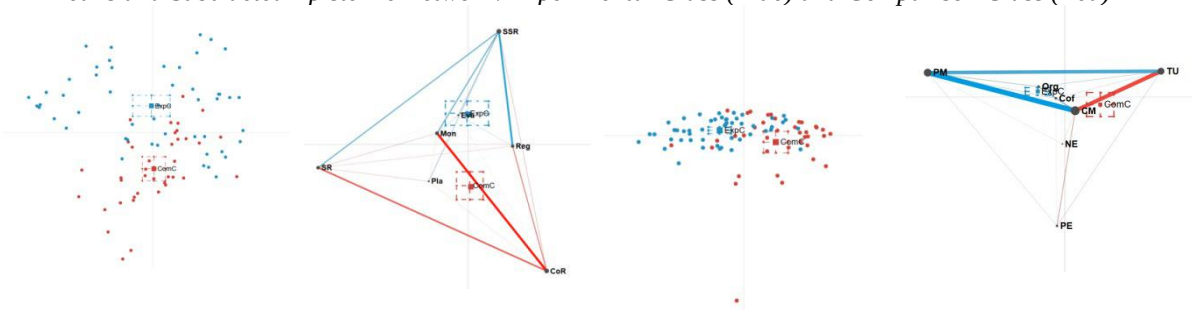
Figure 2 presents the epistemic networks of metacognitive regulation for both classes, illustrating co-occurrences of regulatory behaviors. Regarding regulation type and process, both classes primarily engaged in Socially Shared Regulation (SSR), Co-regulation (CoR), Monitoring (Mon), and Regulating (Reg), with strong connections between Mon and Reg. Conversely, connections involving Self-regulation (SR), Planning (Pla), and Evaluating (Eva) were weaker. Independent-samples t-tests on class centroids showed a significant difference along the y-axis ($t = 7.93$, $p < .001$, $d = 1.70$), indicating stronger community-level regulation in the experimental class. The subtracted network revealed that the experimental class established stronger links between SSR and Pla, Mon,

 © ISLS

Reg, and Eva, demonstrating that students coordinated group activities across the entire inquiry process. Notably, strong connections between Eva and other process nodes suggest the experimental class continuously reflected on outcomes and adjusted future actions, supporting richer, cyclic SSR. By contrast, the comparison class concentrated on CoR or individual monitoring, with weaker reflection and evaluation, resulting in less effective collective regulation.Regarding regulation focus, both classes emphasized Content Monitoring (CM), Task Understanding (TU), and Process Monitoring (PM). However, subtracted network analysis and t-tests revealed that the experimental class prioritized process monitoring (TU-PM) and integrated content and process management (CM-PM), whereas the comparison class primarily focused on content monitoring (TU-CM). The experimental class also exhibited modest links between Organization (Org) and other nodes, reflecting intentional coordination and community responsibility. Overall, the ENA indicates that analytics-supported reflective assessment promoted richer, community-oriented metacognitive regulation in the experimental class, emphasizing the integration of planning, monitoring, evaluation, and regulation, while balancing content and process. In contrast, the comparison class engaged in more superficial, content-focused regulation with weaker reflective cycles.

Figure 2

ENA Means and Subtracted Epistemic Network: Experimental Class (Blue) and Comparison Class (Red)



(a) (b) (c) (d)

Comparisons of regulation type and process for the experimental class (blue) and comparison class (red) using plotted points and centroids (a); Subtracted epistemic network of regulation type and process: the experimental class (blue) and comparison class (red) (b); Comparisons of regulation focus for the experimental class (blue) and comparison class (red) using plotted points and centroids (c); Subtracted epistemic network of regulation focus: the experimental class (blue) and comparison class (red) (d).

Mechanism of analytics-supported reflective assessment for supporting metacognitive regulation (RQ3)

Proportions of metacognitive regulation over time in the experimental class

Table 3 compares the proportion of undergraduates' regulation behaviors in the experimental class before (Stage 1) and during (Stage 2) collaborative analytic-supported reflective assessment. The results showed statistically significant differences between the two stages in all regulation dimensions. In terms of regulation type ($\chi^2(df = 2, N = 1721) = 87.44$, $p < .001$), the experimental class significantly improved SSR (63.28% vs. 40.73%) in Stage2. Regarding the regulation process (χ^2 ($df = 3, N = 1721) = 93.78$, $p < .001$), students in Stage 2 experienced more regulating (42.78% vs. 33.56%) and evaluating (6.15% vs. 0.83%) than in Stage 1. The proportion of planning behaviors gradually decreased. This may be related to the nature of the task, as the early stages of inquiry require more trial-and-error and planning, along with monitoring the group's progress. In the later stage, the focus shifts more towards reflecting on the inquiry process and outcomes. As for the regulation focus ($\chi^2(df = 6, N = 1721) = 45.73$, $p < .001$), students in Stage 2 were more focused on Task Understanding (24.33% vs. 21.2%) and notably Organizing (6.33% vs. 0.5%) compared to Stage 1, and students' expressions of confusion decreased significantly. These results suggested that learning analytics-supported reflective assessment prompted students to focus more on the socially shared regulation of the community, helped them engage in process regulation and evaluation activities, and encouraged a greater focus on organizing group learning tasks in the knowledge building environment.

Epistemic networks of metacognitive regulation over time in the experimental classes

Figure 3 illustrates the epistemic networks of metacognitive regulation in the experimental class across Stage 1 and Stage 2, along with a subtracted network highlighting differences. Red and blue dots indicate individual

centroids for Stage 1 and Stage 2, respectively.Regarding regulation type and process, t-tests showed a significant difference along the y-axis (t = 8.72, p < .001, d = 1.77), but not the x-axis (t = -1.64, p = 0.10). This indicates a shift toward more focused and systematic regulation in Stage 2, where localized centroids suggest clearer and more structured regulatory goals. The subtracted network revealed weakened Co-regulationbut strengthened Socially Shared Regulation, reflecting a shift toward community-level regulation. Additionally, links involving Evaluation (Eva) strengthened while Planning-Monitoring (Pla-Mon) connections decreased, demonstrating that students increasingly integrated reflection and regulation to collaboratively advance knowledge.Regarding regulation focus, t-tests showed no significant differences along either axis (x-axis: t = 0.50, p = 0.62; y-axis: t = -0.81, p = 0.42). Core nodes like Content Monitoring (CM), Task Understanding (TU), and Process Monitoring (PM) remained central across both stages. However, the subtracted network revealed stronger connections involving Organization (Org) in Stage 2, particularly with TU, CM, and PM. This suggests students increasingly applied organizational strategies alongside cognitive behaviors, reflecting more structured coordination. Overall, the analysis indicates that analytics-supported reflective assessment fostered more conscious, systematic, and socially shared metacognitive regulation. Students in Stage 2 integrated monitoring, evaluation, and regulation more effectively, enhancing collaborative engagement within the Knowledge Building environment.

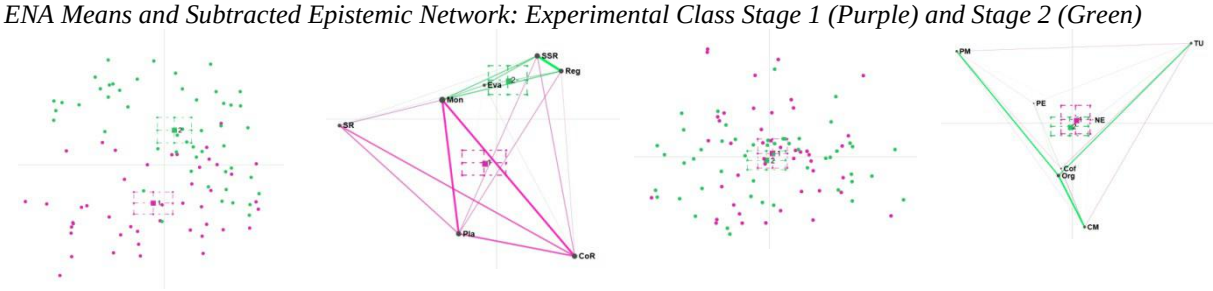
Table 3

Proportions of Metacognitive Regulation of Two Stages in the Experimental Class

	Stage1		Stage2		Total		Change scores from Stage1 to Stage2(%)
	f	%	*f*	%	*f*	%	
Regulation type							
Self-regulation	139	23.21	199	17.74	338	19.64	-5.47
Co-regulation	216	36.06	213	18.98	429	24.93	-17.08
Socially shared regulation	244	40.73	710	63.28	954	55.43	22.55
Regulation process							
Planning	97	16.19	53	4.72	150	8.72	-11.47
Monitoring	296	49.42	520	46.35	816	47.41	-3.07
Regulating	201	33.56	480	42.78	681	39.57	9.22
Evaluating	5	0.83	69	6.15	74	4.30	5.32
Regulation focus							
Task understanding	127	21.20	273	24.33	400	23.24	3.13
Content monitoring	264	44.07	465	41.44	729	42.36	-2.63
Process monitoring	159	26.54	261	23.26	420	24.40	-3.28
Positive emotion	23	3.84	33	2.94	56	3.25	-0.9
Negative emotion	2	0.33	5	0.45	7	0.41	0.12
Confusion	21	3.51	14	1.25	35	2.03	-2.26
Organizing	3	0.50	71	6.33	74	4.30	5.83

Figure 3

ENA Means and Subtracted Epistemic Network: Experimental Class Stage 1 (Purple) and Stage 2 (Green)



(a) (b) (c) (d)

Comparisons of regulation type and process for the experimental class stage1 (purple) and stage2 (green) using plotted points and centroids (a); Subtracted epistemic network of regulation type and process: the experimental class stage1 (purple) and stage2 (green) (b); Comparisons of regulation focus for the experimental class stage1 (purple) and stage2 (green) using plotted points and centroids (c); Subtracted epistemic network of regulation focus: the experimental class stage1 (purple) and stage2 (green) (d).

Discussion and conclusions

Fostering metacognitive regulation through learning analytics-supported reflective assessment

This study demonstrates that learning analytics-supported reflective assessment (LAsRA) significantly enhances undergraduates' collaborative knowledge building and metacognitive regulation (Yang, Feng, et al., 2024). Group Communication Analysis (GCA) reveals that the experimental class significantly outperformed the comparison class in dimensions such as Social Impact and Newness. Such high-cohesion discourse facilitated the integration of heterogeneous perspectives and the implementation of effective Socially Shared Metacognitive Regulation (Shin et al., 2025), while highlighting the core role of students assuming collective cognitive responsibility in advancing knowledge (Cacciamani et al., 2021).

Combining content analysis and Epistemic Network Analysis (ENA), the experimental class exhibited more typical deep-level regulation features: students moved beyond superficial consensus to actively identify conceptual gaps and continuously evaluate and improve ideas (De Backer et al., 2022; Yang et al., 2023). The cross-stage comparison of ENA further revealed the evolution of this mechanism: before the intervention (Stage 1), metacognitive nodes such as planning and evaluating were weakly connected, and regulation relied primarily on preliminary monitoring of outcomes; after the intervention (Stage 2), socially shared regulation was tightly coupled with monitoring and evaluating nodes, successfully constructing a collaborative closed-loop of "goal-setting—process monitoring—reflective evaluation—dynamic regulation". Specifically, the significant increase in the connection strength of the Evaluating node proved to be a necessary condition for producing high-quality task outcomes (Zhang, Chen, et al., 2021).

Furthermore, the notable increase in organizational behaviors in Stage 2 marked the awakening of students' collective epistemic agency (Yang, Chen, et al., 2020). Data feedback from LAsRA empowered the student community to co-configure knowledge-building practices (Tao & Zhang, 2021; Yang, Feng, et al., 2024) and collectively reflect to regulate collaborative inquiry (Zhang et al., 2022). At the affective level, the significant decrease in students' confusion after reflection indicated that visual data not only optimized cognitive strategies but also improved emotional states, corroborating the critical role of socio-emotional regulation in enhancing learning performance (Zhang, Chen, et al., 2021; Yang, Yuan, et al., 2024).

Educational implications

First, fostering a collaborative and reflective culture is essential. Principle-based activities can encourage students to engage in socially shared metacognitive activities, enhancing collective knowledge construction. Second, reflective assessment supports meta-level dialogue, allowing students to monitor, evaluate, and regulate their inquiry effectively. Third, technology-supported analytics can provide real-time, embedded feedback, prompting students to engage in higher-order discussion and reflection. Educators should design learning environments that integrate analytics and structured prompts to scaffold metacognitive development.

Limitations and future research

This study has limitations. First, it did not examine long-term transfer of metacognitive skills to other contexts. Second, the analysis relied primarily on online discourse, lacking multimodal data from face-to-face interactions. Future studies could incorporate AI-based analysis of video, audio, and other process-oriented data. Third, the scope of analytics was limited; future designs could integrate AI for automated, process-oriented reflective interventions across disciplines and learning contexts(Ouyang et al., 2023).

Conclusions

This study demonstrates that learning analytics-supported reflective assessment (LAsRA) significantly enhances undergraduates' collaborative knowledge creation and metacognitive regulation. Through analytics-supported reflective assessment, the intervention prompted students to shift from individual or superficial monitoring toward effective Socially Shared Metacognitive Regulation. Epistemic Network Analysis revealed that LAsRA successfully fostered a dynamic collaborative cycle of "goal-setting — process monitoring — reflective evaluation—dynamic regulation." Within this cycle, students engaged in deep-level regulation by actively identifying gaps and evaluating ideas. These findings provide robust empirical evidence that technology-enhanced reflective assessment establishes a community-level, self-improving regulatory environment, preparing students for continuous innovation in complex collaborative contexts.

 © ISLS

References

Cacciamani, S., Perrucci, V., & Fujita, N. (2021). Promoting students' collective cognitive responsibility through concurrent, embedded and transformative assessment in blended higher education courses. *Technology, Knowledge and Learning*, *26*(4), 1169–1194.

De Backer, L., Van Keer, H., & Valcke, M. (2015). Socially shared metacognitive regulation during reciprocal peer tutoring: Identifying its relationship with students' content processing and transactive discussions. *Instructional Science*, *43*(3), 323–344.

De Backer, L., Van Keer, H., De Smedt, F., Merchie, E., & Valcke, M. (2022). Identifying regulation profiles during computer-supported collaborative learning and examining their relation with students' performance, motivation, and self-efficacy for learning. *Computers & Education*, *179*, 104421.

D'Mello, S., Lehman, B., Pekrun, R., & Graesser, A. (2014). Confusion can be beneficial for learning. *Learning and Instruction*, *29*, 153–170.

Dowell, N. M., Nixon, T. M., & Graesser, A. C. (2019). Group communication analysis: A computational linguistics approach for detecting sociocognitive roles in multiparty interactions. *Behavior Research Methods*, *51*(3), 1007–1041.

Dowell, N. M., & Poquet, O. (2021). SCIP: Combining group communication and interpersonal positioning to identify emergent roles in scaled digital environments. *Computers in Human Behavior*, *119*, 106709.

Hadwin, A. F., Järvelä, S., & Miller, M. (2011). Self-regulated, co-regulated, and socially shared regulation of learning. In B. J. Zimmerman & D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 65–84). Routledge.

Järvelä, S., & Hadwin, A. F. (2013). New frontiers: Regulating learning in CSCL. *Educational Psychologist*, *48*(1), 25–39.

Järvelä, S., Kirschner, P. A., Hadwin, A., Järvenoja, H., Malmberg, J., Miller, M., & Laru, J. (2016). Socially shared regulation of learning in CSCL: Understanding and prompting individual- and group-level shared regulatory activities. *International Journal of Computer-Supported Collaborative Learning*, *11*(3), 263–280.

Järvelä, S., Kirschner, P. A., Panadero, E., Malmberg, J., Phielix, C., Jaspers, J., & Järvenoja, H. (2015). Enhancing socially shared regulation in collaborative learning groups: Designing for CSCL regulation tools. *Educational Technology Research and Development*, *63*(1), 125–142.

Kim, D., & Lim, C. (2018). Promoting socially shared metacognitive regulation in collaborative project-based learning: A framework for the design of structured guidance. *Teaching in Higher Education*, *23*(2), 194–211.

Lee, A., O'Donnell, A. M., & Rogat, T. K. (2015). Exploration of the cognitive regulatory sub-processes employed by groups characterized by socially shared and other-regulation in a CSCL context. *Computers in Human Behavior*, *52*, 617–627.

Lei, C., & Chan, C. K. K. (2018). Developing meta-discourse through reflective assessment in knowledge building environments. *Computers & Education*, *126*, 153–169.

Miller, M., & Hadwin, A. (2015). Scripting and awareness tools for regulating collaborative learning: Changing the landscape of support in CSCL. *Computers in Human Behavior*, *52*, 573–588.

Oshima, J., Oshima, R., & Matsuzawa, Y. (2012). Knowledge building discourse explorer: A social network analysis application for knowledge building discourse. *Educational Technology Research and Development*, *60*(5), 903–921.

Ouyang, F., Xu, W., & Cukurova, M. (2023). An artificial intelligence-driven learning analytics method to examine the collaborative problem-solving process from the complex adaptive systems perspective. *International Journal of Computer-Supported Collaborative Learning*, *18*(1), 39–66.

Panadero, E., & Järvelä, S. (2015). Socially shared regulation of learning: A review. *European Psychologist*, *20*(3), 190–203.

Pekrun, R., Vogl, E., Muis, K. R., & Sinatra, G. M. (2017). Measuring emotions during epistemic activities: The epistemically-related emotion scales. *Cognition and Emotion*, *31*(6), 1268–1276.

Raes, A., Schellens, T., De Wever, B., & Benoit, D. F. (2016). Promoting metacognitive regulation through collaborative problem solving on the web: When scripting does not work. *Computers in Human Behavior*, *58*, 325–342.

Scardamalia, M. (2002). Collective cognitive responsibility for the advancement of knowledge. In B. Smith (Ed.), *Liberal education in a knowledge society* (pp. 67–98). Open Court.

Scardamalia, M., & Bereiter, C. (2006). Knowledge building: Theory, pedagogy, and technology. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 97–118). Cambridge University Press.

Scardamalia, M., & Bereiter, C. (2014). Knowledge building and knowledge creation: Theory, pedagogy, and technology. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (2nd ed., pp. 397–417). Cambridge University Press.

Shin, Y., Lee, H., & Kim, W. (2025). Group cohesiveness and socially shared metacognitive regulation. *Thinking Skills and Creativity*, *56*, 101771.

Siemens, G., & Gašević, D. (2012). Guest editorial-learning and knowledge analytics. *Journal of Educational Technology & Society*, *15*(3), 1–2.

Snell, J., & Lefstein, A. (2018). "Low ability," participation, and identity in dialogic pedagogy. *American Educational Research Journal*, *55*(1), 40–78.

Su, Y., Li, Y., Hu, H., & Rosé, C. P. (2018). Exploring college English language learners' self and social regulation of learning during wiki-supported collaborative reading activities. *International Journal of Computer-Supported Collaborative Learning*, *13*(1), 35–60.

Tao, D., & Zhang, J. (2021). Agency to transform: How did a grade 5 community co-configure dynamic knowledge building practices in a yearlong science inquiry? *International Journal of Computer-Supported Collaborative Learning*, *16*, 403–434.

Volet, S., Vauras, M., & Salonen, P. (2009). Self- and social regulation in learning contexts: An integrative perspective. *Educational Psychologist*, *44*(4), 215–226.

White, B., & Frederiksen, J. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Instruction*, *16*(1), 3–118.

Yang, Y., Chen, Q., Yu, Y., Feng, X., & van Aalst, J. (2020). Collective reflective assessment for shared epistemic agency by undergraduates in knowledge building. *British Journal of Educational Technology*, *51*(4), 1136–1154.

Yang, Y., van Aalst, J., & Chan, C. K. K. (2020). Dynamics of reflective assessment and knowledge building for academically low-achieving students. *American Educational Research Journal*, *57*(3), 1241–1289.

Yang, Y., van Aalst, J., Chan, C. K. K., & Tian, W. (2016). Reflective assessment in knowledge building by students with low academic achievement. *International Journal of Computer-Supported Collaborative Learning*, *11*, 281–311.

Yang, Y., Zhu, G., Sun, D., & Chan, C. K. K. (2022). Collaborative analytics-supported reflective assessment for scaffolding pre-service teachers' collaborative inquiry and knowledge building. *International Journal of Computer-Supported Collaborative Learning*, *17*, 249–292.

Yang, Y., Feng, X., Zhu, G., & Sun, D. (2023). Exploring the mechanisms of data-supported reflective assessment for pre-service teachers' knowledge building. *Interactive Learning Environments*, *32*(9), 5660–5677.

Yang, Y., Yuan, K., Zhu, G., & Jiao, L. (2024). Collaborative analytics-enhanced reflective assessment to foster conducive epistemic emotions in knowledge building. *Computers & Education*, *209*, 104950.

Yang, Y., Feng, X., Zhu, G., & Xie, K. (2024). Effects and mechanisms of analytics-assisted reflective assessment in fostering undergraduates' collective epistemic agency in computer-supported collaborative inquiry. *Journal of Computer Assisted Learning*, *40*(3), 1098–1122.

Zhang, J., Tao, D., Chen, M. H., Sun, Y., Judson, D., & Naqvi, S. (2018). Co-organizing the collective journey of inquiry with idea thread mapper. *Journal of the Learning Sciences*, *27*(3), 390–430.

Zhang, J., Tian, Y., Yuan, G., & Tao, D. (2022). Epistemic agency for costructuring expansive knowledge-building practices. *Science Education*, *106*(4), 890–923.

Zhang, S., Chen, J., Wen, Y., Chen, H., Gao, Q., & Wang, Q. (2021). Capturing regulatory patterns in online collaborative learning: A network analytic approach. *International Journal of Computer-Supported Collaborative Learning*, *16*(1), 37–66.

Zhang, S., Wen, Y., & Liu, Q. (2021). Exploring student teachers' social knowledge construction behaviors and collective agency in an online collaborative learning environment. *Interactive Learning Environments*, *29*(4), 543–556.

Zheng, J., Xing, W., & Zhu, G. (2019). Examining sequential patterns of self- and socially shared regulation of STEM learning in a CSCL environment. *Computers & Education*, *136*, 34–48.

International Society of the Learning Sciences

Extending CSCL Affordances for Multilingual Classrooms: Teacher Perspectives on AI-Supported Collaboration

Pragati Maheshwary, Alina Guha, Shamya Karumbaiah
pmaheshwary@wisc.edu, ajguha@wisc.edu, shamya.karumbaiah@wisc.edu
University of Wisconsin-Madison

Abstract: Multilingual classrooms add layers of complexity to collaboration: teachers rarely share every student's home language, peers struggle to build common ground, and AI translation can miss linguistics nuance. We report findings from focus groups conducted with teachers of multilingual students to gauge their view on AI-supported collaboration across three collaboration pathways: student-student, teacher-student, and student-AI. Teachers welcomed real-time subtitling/translation to support comprehension and questioning, and lightweight tools that surface peers' ideas via transcripts and highlights for clarifications. They were skeptical of in-class 1:1 student-AI partnering, citing risks to social connection, teacher workload, and scope creep. Using Jeong & Hmelo-Silver's seven affordances as our lens, we identify where affordances blend within pathways, and where they need rethinking for multilingual equity. We translate teacher concerns into CSCL-aligned design adaptations and contribute extended versions of affordances to guide AI-integrated CSCL that is equitable for linguistically diverse classrooms.

Introduction

Collaborative learning positions knowledge building as a social process in which learners co-construct understanding through discourse, shared artifacts, and interdependence (Smith & MacGregor, 1992). Decades of Computer Supported Collaborative Learning (CSCL) research demonstrate that well-orchestrated collaboration supported with appropriate tools, scaffolds, and assessment improve achievement and engagement across domains (Chen et al., 2018).

In our work, we conceptualize collaboration as manifesting along three complementary interactional pathways: (1) *student-student collaboration,* in the form of small group discussions, problem solving, and knowledge co-construction, (2) *teacher-student collaboration,* in the form of teachers engaging with and encouraging student ideas, teachers providing conceptual guidance and resources, and students asking clarifying questions, and (3) emerging *student-AI collaborations,* in the form of students engaging with agents, robots or chatbots for self-regulated learning. Student-student collaboration leverages peer explanation and shared regulation to deepen reasoning (Järvelä et al, 2018). Teacher-student exchanges remain foundational for framing tasks, diagnosing understanding, and apprenticing disciplinary talk (Erkan, 2019). Finally, as conversational agents enter classrooms, there is growing interest in the role of generative AI as a participant that may facilitate brainstorming, feedback, orchestration and knowledge co-construction with human learners (Benson, 2023; Wei et al., 2025; Choi et al., 2025).

Nevertheless, these collaborations are hard to enact equitably in multilingual classrooms. Teachers often do not share students' home languages; peers may not understand one another; and tools that translate words can still miss cultural nuance, dialectal variation, and the multimodal ways in which students "language" ideas (Barron et. al., 2025). Teachers worry about over-reliance on translation, amplification of safety risks in low-resource languages, and the practical burden of monitoring many groups simultaneously (Guha et al., 2025a). At the same time, they see promise in valuing students' full linguistic repertoires through strategic use of their home and classroom languages, visuals, and multimodal resources, if supported by technology and human mediation to preserve pedagogical intent (Guha et al., 2025b).

CSCL literature provides several useful frameworks for addressing these tensions but the intersection of AI-integrated CSCL and multilingual settings remains underexplored. Early studies suggest AI can influence team processes and creativity, but they also underscore risks to equity, agency, and discourse quality if designs are not carefully constrained (Wei et al., 2025). We therefore adopt Jeong & Hmelo-Silver's (2016) seven affordances of CSCL as our analytic lens, which articulates how technology can enable and shape productive collaboration. The seven interdependent affordances defined by Jeong & Hmelo-Silver that together capture the social, cognitive, and technological mechanisms of collaborative learning are: (1) *Establishing and understanding joint tasks* (i.e., ensuring learners share goals and awareness of the activity), (2) *Communication* (i.e., supporting the exchange of ideas and questions), (3) *Resource sharing* (i.e., enabling access to information, representations, and artifacts), (4)

 © ISLS

Engaging in productive processes (i.e., facilitating collective reasoning, elaboration, and reflection), (5) *Engaging in co-construction* (i.e., coordinating meaning-making and knowledge building), (6) *Monitoring/regulation* (i.e., sustaining shared awareness and self-regulation across participants), and (7) *Finding and building community* (i.e. fostering sustained relationships and collaboration networks). We adopted this framework because it is technology-agnostic, spans specific mechanisms of collaborative learning (e.g., discourse moves, shared artifacts, regulation), and gives us a principled way to trace how technology can support collaboration across our three collaboration pathways: teacher-student, student-student, and student-AI interactions. Building on prior CSCL literature, this framework provides a structured lens to examine and extend CSCL affordances for multilingual contexts. We also drew on Wise and Schwarz's (2017) provocations for the future of CSCL as a conceptual backdrop when framing our study and interpreting its implications. Their provocations challenge the field to reimagine how collaborative learning research evolves amid shifting technologies and sociocultural conditions. In particular, their calls to re-center learner agency (Provocation #2), emphasize contextual adaptation and alignment with real-world practices (Provocation #4), and reaffirm teacher-mediated orchestration as the anchor of technological innovation (Provocation #8) guided our conceptual orientation.

Partnering with teachers who navigate local curricula and students' linguistic/cultural histories in multilingual classrooms, in this paper, we examine AI's role across teacher-student, student-student, and student-AI collaborations. We synthesize teachers' views on the appropriate use of AI, evaluate how Jeong & Hmelo-Silver's seven affordances manifest in teachers' concerns and extend them where reinterpretation is required. Finally, we surface concrete design adaptations that make CSCL practices linguistically inclusive by allowing students access to their home language in collaborative learning, while supporting teacher oversight on safe use of AI and students' development of classroom language. In summary, we show how the original CSCL affordances from Jeong & Hmelo-Silver need to be extended for linguistically diverse classrooms.

Methods

Building on a survey study conducted by Guha et. al., 2025a, Guha et. al. 2025b with teachers of multilingual learners, we identified real-world representative scenarios of multilingual classrooms that may struggle with the three mentioned collaboration pathways. Taking inspiration from teacher responses to open-ended questions, we also developed potential AI solutions as a point of deliberation for further discussion on AI affordances and concerns regarding AI support for collaboration in multilingual classrooms. For student group collaboration, we presented a tool that can provide translations during small group discussions supported by AI. Students can use the tool to highlight comments from peers for clarification and teachers can view the students' conversation transcripts to ask the tool questions about what was discussed (*Tool #1*). For student-teacher collaboration supported by AI, we presented a tool that can provide all students real-time translations of teacher's speech during a whole group lecture and the reverse translation of student speech to teachers for when the student would like to ask a clarifying question (*Tool #2*). For student-AI-teacher collaboration, we presented a 1:1 AI conversational partner that can support students in understanding a concept and communicating in the language of instruction; the teacher can later view the transcript and understand where the student may be struggling (*Tool #3*).

The study was IRB approved and recruitment was conducted by emailing teachers who participated in the initial survey study. We conducted 4 focus groups with 15 teachers all of whom taught multilingual students. Each focus group was 60 minutes and consisted of 3-4 teachers. Each scenario and solution was presented through google slides with some guiding questions and discussed by at least 2 groups of teachers. 13 of these teachers were teaching in the United States, the other 2 were in Brazil and Indonesia respectively. 6 teachers were English support teachers (e.g. English as Second Language (ESL), while the others taught content areas such as Science, Social Studies, Math, English, Kichwa language studies, Life Sciences, and Information Technology. There was a balance across grade levels such that we had 4 high school teachers, 3 middle school teachers, 3 kindergarten/elementary school teachers and 2 teaching assistants at the post-secondary level.

To analyze the focus group data, we followed the affinity diagramming method as described by Lawrence and colleagues (2023). We analyzed 119 data points and inductively clustered feedback on presented scenarios into concerns and design adaptations. To achieve reliability, two researchers analyzed the focus group transcripts from which data unique quotes were extracted as data points. Two additional researchers analyzed data points negotiating themes and reviewing the data until consensus was reached. Through an iterative process of clustering, two researchers leveraged those themes to identify both unique and common concerns and design ideas for each collaboration type.

Results

We frame our findings using the seven CSCL affordances proposed by Jeong & Hmelo-Silver (2016) after extending it to use of AI to support collaborative learning in multilingual and multicultural classroom contexts

as: (1) establishing understanding of collaborative tasks, (2) communicating multilingually, (3) sharing resources in students' home and classroom language, (4) structuring language-scaffolded productive processes, (5) facilitating co-construction in students' home and classroom language, (6) monitoring/regulation, and (7) finding and building inclusive groups and communities. While Jeong & Hmelo-Silver (2016)'s CSCL affordances describe what collaboration technologies should enable, our focus groups revealed how those affordances must be re-interpreted in multilingual, AI-mediated classrooms from teachers' standpoint. We further describe how these newly defined affordances show-up along three collaboration pathways: student-student, teacher-student, and student-AI-teacher, showing how teachers envision AI not as a replacement for human roles but as a supporting tool that scaffolds comprehension, connection, and linguistic growth.

Student group collaboration supported by AI

Jeong & Hmelo-Silver's framework begins with the affordance of communication, which students build on to engage in productive processes and co-construction of knowledge. This is followed by students' ability to monitor/regulate their own collaborative process. Teachers described small-group dialogue as an essential practice in multilingual classrooms to support peer learning and language growth yet challenging to put into practice. One teacher noted this by saying, *"we don't all have a common language, we can't really converse [by] speaking, and so I think that's our biggest barrier"*. Therefore, they welcomed the idea of a transcripts-and-highlighting tool (*Tool #1*; see Methods section) to facilitate discussions and understanding across languages.

Table 1

Opportunities/Challenges and Design Adaptations that Surfaced in Focus Group Discussion on Transcripts-and-Highlighting Tool

Student-Student Collaboration Supported by AI			
Opportunities/Challenges to be Addressed	*Illustrated Quotes*	*Proposed Design Adaptations*	*Extended CSCL Affordance*
Minimal adult intervention to support peer-peer connection	*provide language frames, and language support in English so they can start, communicating with their peers, right, with pieces here and there*	Built-in peer-facilitation features that prompt students to rotate roles (facilitator/scribe), turn-taking queue, and sentence starters in home & classroom language	Communicating multilingually + Structuring language-scaffolded productive processes
	if an adult has to step in and translate, and then you're not getting that peer-to-peer connection, and so it becomes more difficult	Clarify or "I didn't get that" button for peers to request restatement without teacher involvement	Facilitating co-construction in students' home and classroom language
Ability to track on-task vs off-task conversations	*you're not exactly sure, like, what is being said [in the different language], if they're on track, if they're off track,*	Teacher notifications for off-task conversations to bring back teachers in the loop	Structuring language-scaffolded productive processes
		Subtle on-topic nudge to students when drift is detected	Structuring language-scaffolded productive processes
Reliability in knowing who said what to promote shared-regulation	*the reliability of knowing who it was that was saying it. Like, even if it's going into their device from their conversation, how is this program making sure it was Susan and not, you know, Stanley or whatever.*	Automated speaker recognition with color labels and a tap-to-relabel feature if wrong for transcripts	Monitoring/Regulation

Teachers insisted that the tool structure productive processes through built-in facilitation features like role rotation, a turn-taking queue, and sentence starters in student's home & classroom language that minimize the need for adult intervention and increase peer-peer connection. With the aforementioned features we extend the *engaging in productive processes* affordance to *structuring language-scaffolded productive processes*. This helps students negotiate meaning across languages and AI acts as an equity facilitator, not dictating discourse but supporting distributed understanding. Promoting learner agency by providing a clarification button to seek more

information from peers across languages further promotes *co-construction of knowledge in students' home and classroom languages* in a multilingual context. Teachers also expressed concerns about students getting distracted in small group discussions and not being able to track off-topic conversations, so they desired more oversight. Teacher notifications or student nudges can provide light-touch oversight on off-task behaviors thereby further *structuring language-scaffolded productive processes*. Lastly, allowing the students to see an accurate representation of their conversation as it unfolds and after the conversation ends affords them the ability to *monitor/regulate* their own collaborative processes independently. While the AI tool can automatically detect the speakers (aka speaker diarization), allowing students to relabel instances of inaccurate speaker detection can improve AI reliability and student agency. Table 1 details how teachers' needs for minimal adult mediation, ability to track off-task behavior and reliability yield concrete scaffolds that add equity-by-design to Jeong & Hmelo-Silver's affordances.

Teacher-student collaboration supported by AI

Jeong & Hmelo-Silver mention communication and sharing resources as essential starting points for the collaborative learning process. Teachers, however, find communicating with students and sharing appropriate resources in a multilingual classroom difficult as students may come from diverse linguistic backgrounds unfamiliar to the teachers. Hence, teachers welcomed the idea of a real-time subtitling/translation tool (*Tool #2*;see Methods section) to enable students to listen in English while receiving support to comprehend the content and ask questions about it in their home language, redefining the communication affordance as *communicating multilingually*. Teachers suggested that progressively fading the home language support through the translation tool alongside can help overcome reliance on AI and promote English Language Acquisition.

Table 2

Opportunities/Challenges and Design Adaptations that Surfaced in Focus Group Discussion on real-time subtitling/translation tool

Teacher-Student Collaboration supported with AI			
Opportunities/Challeng -es to be Addressed	*Illustrated Quotes*	*Proposed Design Adaptations*	*Extended CSCL Affordances*
Overcome language barrier to support content learning	*so that they can still access that content, but in a more visual way as they can associate what they're, what they're listening to*	Add images alongside home text to establish associations	Sharing resources in students' home and classroom language
	…root words in their native language that are similar to those in English to try to make the connection… and use both sides of their language brain.	Recommend root words/cognate cues from students' home languages that are similar to those in English to try to make the connection	Facilitating co-construction in students' home and classroom language
Over-reliance on AI could stall English language acquisition	*helping build them that, um, English vocabulary through some other scaffolds [like] breaking down more complex words, giving, um, more basic English definitions*	Provide teacher recommendations to scaffold english itself by breaking down complex words and providing more basic english definitions; Implement practice nudges like retrieval prompts, spacing, and quick English pronunciation aids; Provide teachers the ability to start and pause student's access to transcripts and other features of the tool to progressively fade translation support;	Communicating multilingually
Cognitive Overload	*having a student have to listen to the English but then also read in their home language that mental lift is quite exhausting*	Use the tool only for 1:1 interactions instead of using it during classroom discussions to avoid student disengagement due to juggling between listening in one language while reading in another	Structuring language-scaffolded productive processes

Teachers requested that the tool provide teachers support for engaging students in creative ways by providing recommendations on the use of simpler definitions, visuals and cognate cues, thus extending the affordance to *sharing resources in students' home and classroom languages*. Using these modifications rather than direct translation allows students to access the content through the lens of their own experiences. For example, using an image can help students relate the content the teacher is speaking about to their home life even if they may not have the vocabulary of the concept in their home or classroom languages. Lastly, teachers wanted more control on constraining when or how the tool is used, specifically, using it only for 1:1 interaction to manage students' divided attention caused due to potential cognitive overload from juggling between listening in one language while reading in another. Table 2 summarizes how concerns about language acquisition, concept learning, and cognitive load translate into design adaptations that extend these affordances in practice.

Table 3

Opportunities/Challenges and Design Adaptations that Surfaced in Focus Group Discussion on 1:1 Personalized AI Support

Student-AI-Teacher Collaboration			
Opportunities/Challeng -es to be Addressed	*Illustrated Quotes*	*Proposed Design Adaptations*	*Extended CSCL Affordances*
Maintaining in-class engagement	*I really like the idea of flipped classrooms, where we're using [edtech] videos and then using that translation setting to have, like a library of videos for each, big-idea content, objective we're teaching.*	Utilize the tools in a flipped classroom structure so that students go through content, seek clarification from the AI and bring to class more questions for the teacher and peers	Co-construction in students' home and classroom language
Provide personalized support while limiting the scope of AI's response to the discussion topic	*tell the AI chatbot, like, "Please keep the discussion to the topic and kind of take out that extraneous details?" I wanna make sure my students are getting proper information and aren't getting this, like, maybe skewed based on what is posted mostly on the internet*	Tool should not introduce new concepts but only help with current content understanding through a variety of materials like videos, and diagrams; if questions on new content come up there should be "ask-teacher" flags	Sharing resources in students' home and classroom language
Added labor for teachers to oversee student-AI collaboration	*you can view a summary of this conversation. You know, students can use this under your supervision. I'm not sure if teachers will be willing to do it because that's technically extra work for teachers other than teaching, because teaching is already very, very tiring*	Tool can provide short bulleted summaries with highlights for misconceptions and attached evidence quotes from transcripts	Structuring language-scaffolded productive processes
Isolation leading to reduced connection with teachers and reduced interactions with peer	*[students] might limit their interactions with the teacher, so I wonder if the AI could be trained to develop their English proficiency...so then, in some ways it could be a means to having a greater connection*	Tool should provide language support and encourage students to ask questions to the teacher and peers	Finding and building inclusive groups and communities + Communicating multilingually

Student-AI-teacher collaboration

Jeong & Hmelo-Silver's framework goes beyond supporting productive processes, sharing resources, and facilitating co-construction to also assist students in independently finding and building groups and communities for future collaboration. This framework originally assumed collaboration among students or students and agents/robots, however, based on teacher interactions our final collaboration pathway extends this idea, considering the potential of generative AI as becoming an intermediary participant between teacher and student collaboration. Teachers in our focus groups were skeptical of 1:1 conversational AI that assists their students asynchronously and report to teachers post hoc (*Tool #3*; see Methods section). The tool was meant to introduce

the possibility of personalized support in multilingual classrooms where it may be difficult for teachers to give students individual attention during class time.

Teachers highlighted new demands when AI becomes an intermediary participant between student-teacher collaboration. They endorse targeted 1:1 support by AI only if it (a) maintains in-class engagement and student-student collaboration, (b) constrains content scope (clarifies misconceptions without teaching new content), (c) reports back to the teacher succinctly with evidence quotes, and (d) protects social relationships (avoids isolation from teacher/peers). The student-AI-teacher scenario thus reframes AI as a *teacher-bounded participant* in student learning. While AI directly engages with students through 1:1 interactions, its role is shaped and constrained by teacher-defined expectations. Design adaptations such as summaries, "ask-teacher" flags, controlled content scope, and restricting AI use to flipped classroom settings (Table 3) bound how AI participates—limiting what it can introduce, when it should guide students to the teacher, and how it maintains visibility for teacher oversight. In this way, AI collaborates with students, without displacing teacher authority or peer interaction, instead supporting and channeling student engagement within structured limits.

Common concerns with AI across all collaboration types

While Jeong & Hmelo-Silver's (2016) seven affordances offer a productive lens for analyzing collaboration, teachers' reflections in our study revealed three recurring concern themes common across all collaboration pathways, emphasizing that the technological efficiency/advancement of the "computer-support" layer of CSCL must evolve alongside its pedagogical principles (see Table 4).

Table 4

Common Key Concerns and Design Adaptations that Surfaced Across All Collaboration Types Organized into Major Themes

Commonalities Across All Collaboration Types		
Theme	*Opportunities/Challenges to be Addressed*	*Proposed Design Adaptations*
Socio-cultural and linguistic nuance of multilingual classrooms	Inaccuracies in interpreting low-resource languages (LRL), accents, slangs, dialects, and students with speech disabilities	Prefer models audited on target dialects/LRLs; allow teacher lexicon overrides and show an evaluation card documenting dialect/LRL performances
	Lack of academic vocabulary in home language	Tool should support multilingual input and output as some academic vocabulary translated in students' home language might not make sense or are inaccessible to students
		Provide concept-level explanations with teacher approved glossaries only
	Students might not know how to read in their home language and only know their home language verbally	Tool should allow for voice input and output
		Utilize more quick emoji and visual responses
Pedagogical Trust and Control	Content inaccuracies and hallucinations in an unfamiliar language	Allow teachers to create knowledge base for their course within the tools
	Mentions of inappropriate words from unfamiliar language going undetected	Provide agency to teachers to censor words and review translations before showing to students; teacher-curated terms should override model choices
Managing Developmental Fit	Low technical skills of younger students and limited screen time	Limit the number of features, simplify UI, and place attention guards like timeboxing and focus mode that hides subtitles during key talk
	High school students prefer communicating through text over speech	AI can increase comfort in communication by allowing texting with their peers and providing a safe practice space to later verbally engage in groups or classroom

First, the socio-cultural and linguistic nuance of multilingual classrooms highlight that AI must become linguistically adaptable and safe for multilingual classrooms before it can effectively support collaboration. Teachers pointed to the need for models that show higher accuracy in detecting different accents and low-resource

languages (LRLs), allow overrides from teachers for use of certain words, and allow multilingual input/output that preserve students' intent rather than losing context through literal translation. Teachers also called for multimodal interactions using voice input, emoji, and visual scaffolds to make AI tools accessible to learners with speech differences or literacy gaps. Put together, teachers desired technologies that more accurately reflect multilingual students' diverse ideas through linguistic and cultural nuance, along with multimodality.

Second, pedagogical trust and control, underscored that integrating AI into classrooms requires looping teachers into every step of mediation. Teachers sought greater agency over AI systems through controls to curate glossaries, review or censor content, and access evidence-bearing summaries and model transparency logs. This layer of teacher oversight, anchored in orchestration, ensures that AI augmentation aligns with curricular goals rather than competing with them.

Finally, managing developmental fit called attention to differences in how younger versus older learners interact with technology. Teachers stated that younger students may benefit from simplified interfaces, limited features, and focus guards that reduce distraction, while older students might engage more confidently through text-based rehearsal spaces before speaking. Together, these themes call for rethinking how AI supports collaboration through culturally responsive, teacher-mediated, and developmentally attuned design.

In summary, across the three collaboration pathways i.e., teacher-student, student-student, and student-AI-teacher interactions, we found that teachers believe AI can best support multilingual collaboration when it is responsive to the cultural and linguistic nuances of multilingual classrooms, keeps teachers in control of how it is used, and adapts to learners' developmental needs. Together, these refinements position our results as a practical re-specification of Jeong & Hmelo-Silver's framework for the multilingual, AI-enabled era of collaborative learning.

Discussion

Our findings on teacher perceptions of AI support for collaborative learning in multilingual classrooms resonate with core CSCL affordances as proposed by Jeong & Hmelo-Silver (2016), while extending them to the realities of multilingual classrooms. Research in collaborative learning shows that outcomes improve when roles, scripts, shared artifacts, and assessment are intentionally designed (Chen et al., 2018; Miller & Hadwin, 2015; Smith & MacGregor, 1992). Our proposed design adaptations like teacher-facing controls, role rotations, and context-aware knowledge bases echo these results. Additionally, features recommendations such as curriculum-locked corpora and teacher-curated glossaries extend existing affordances as defined by Jeong & Hmelo-Silver (2016) into multilingual contexts where both communication and comprehension must be mediated across languages.

In our study, we consistently treated collaboration as having three facets: student-student, teacher-student and student-AI, but the teachers stressed that the student-AI case cannot stand alone. The need for teacher oversight is consistent with prior work that emphasizes that orchestration, monitoring, and adaptive guidance by teachers remain central to productive collaborative learning, particularly in technology-mediated environments (Holstein et al., 2019; van Leeuwen, 2015). Therefore, even though the initial scenario imagined an emerging student-AI partnership, teachers asked for human oversight, converting it into student-AI-teacher collaboration, modifying AI's role to become a *teacher-bounded participant* (instead of just a tool or an autonomous participant). Their reasons were practical and relational: keeping classroom relationships intact, preventing scope drift into new content, and ensuring any help returned to the teacher with short evidence quotes and an invitation for follow-up. This preference does not reject AI support but situates it. Teachers also favored targeted, time-boxed use of AI that prepares learners to engage with peers without AI presence but with clearer questions and fewer misconceptions.

In line with the Extended CSCL Affordances in Table 1, 2, and 3, teachers consistently emphasized and repeatedly brought up topics that mapped to certain affordances like: *communicating multilingually* (in-the-moment translation and clarify moves), *structuring language-scaffolded productive processes* (role rotation, turn-taking queues, sentence starters, and light on-task nudges), *facilitating co-construction* (transcripts that show who said what so peers can build shared understanding), *sharing resources in students' home and classroom language* (simpler definitions, visuals, cognate cues). However, affordances such as *establishing understanding of collaborative tasks*, and *finding and building inclusive groups and communities* appeared less often in teachers' reasoning around the different scenarios. This reflects immediate priorities around language access, comprehension, and orchestration control. Evaluating whether the narrowed focus on aforementioned immediate priorities strengthens or limits equitable collaboration requires further empirical work. Including students' perspectives and classroom enactments in empirical studies may surface needs for affordances teachers did not explicitly identify. Including students alongside teachers in the co-design process can also be helpful for understanding whether and how underrepresented affordances like finding and building communities or establishing collaborative tasks can be cultivated as classrooms mature in their use of AI.

Additionally, this work concretely responds to several provocations articulated by Wise and Schwarz (2017) in "Visions of CSCL: eight provocations for the future of the field". In addressing, Provocation #2 (prioritizing learner agency over collaborative scripting), teachers consistently rejected heavy scripting (i.e. associating groups of learners with specific tasks, roles, or resources, and/or constraining the mode of interaction among peers) or fully automated scaffolding, preferring "scaffold-then-fade" structures that respect both learner initiative and teacher judgment. Across collaboration types, the principle of design for guided autonomy was the same. In student-student collaboration, this took the form of light peer facilitation features such as turn-taking queues, role rotation, and "clarify" buttons, that preserve conversational choice while ensuring mutual understanding across languages. In teacher-student interactions, 1:1 translation support and language scaffold recommendations were thought to help comprehension without replacing human dialogue. Lastly, in student-AI personalized support, use of AI in flipped classroom settings and "ask-teacher" flags protect classroom relationships from over-automation. Together, these adaptations reconcile agency and structure through co-regulation rather than control, advancing Wise and Schwarz's vision of learner-centered yet coordinated collaboration.

Our results also respond to Provocation #6 (learning analytics and adaptive support) by grounding adaptation in *teacher mediation*. Teachers in our study wanted analytics that guide, not replace their professional judgment. Mechanisms like evidence-bearing summaries with quote snippets from student-student or student-AI discussions, and adjustable fading schedules for translation, redefine adaptivity as a process of pedagogical alignment, ensuring that what adapts, when, and how remains under human oversight. The shift from generic personalization to instructionally aligned adaptation turns Wise and Schwarz's abstract call for "adaptive support" into a concrete, teacher-controlled reality. The convergence of these responses underscores that CSCL frameworks can be constructed bottom-up, grounded in knowledge of what teachers actually do to manage learning rather than from broad, top-down frameworks.

Finally, our results connect directly to Provocation #8 (CSCL should give up on educational change), not by abandoning improvement, but by relocating change to *orchestration-in-practice*. Rather than promising sweeping reform, teachers in our study mention small, durable moves that travel across teacher-student, student-student, and student-AI-teacher contexts like curriculum-locked corpora, teacher-curated glossaries, evidence-bearing summaries, 1:1 constrained use of in-class translation tool, timeboxing, and show/hide/approve controls. In this view, equity is produced by defaults and routines, not declarations, through linguistically-aware workflows and progressive fading of AI support that value home languages for sense-making while safeguarding long-term growth in the language of instruction. This is changing at the scale of daily practice: credible, sustainable, and owned by teachers.

In conclusion, the implications of this work extend to two communities. Learning designers should work with teachers and students to embed design adaptations mentioned here into different collaboration contexts. Researchers in natural language processing and speech technology must simultaneously advance models along the dimensions teachers identified as essential: stronger support for low-resource languages and dialects; speech systems that work well with many accents and with students who have speech disabilities; explanations that focus on ideas rather than word-for-word translation; and clear indicators when the system is unsure. Progress on both fronts is required to make CSCL's promise credible and equitable in linguistically diverse classrooms. In doing so, we not only extend Jeong & Hmelo-Silver's framework into multilingual, AI-mediated contexts but also materialize Wise & Schwarz's provocations, showing how unity, agency, and adaptive support can be achieved through situated, teacher-driven design.

References

Barron, K., Tseng, N., Karumbaiah, S., & Baeza, C. T. (2025). Evaluating Language Models Using Linguistic Variations in Multilingual Learners' Writing: A Teacher Study. Proceedings of the 19th International Conference of the Learning Sciences-ICLS 2025, Pp. 1389-1393. https://repository.isls.org/handle/1/11281

Benson, A. (2023). The Future of AI in Education: AI Classroom Partners. XRDS: Crossroads, The ACM Magazine for Students, 29(3), 30–35. https://doi.org/10.1145/3589646

Chen, J., Wang, M., Kirschner, P. A., & Tsai, C.-C. (2018). The Role of Collaboration, Computer Use, Learning Environments, and Supporting Strategies in CSCL: A Meta-Analysis. Review of Educational Research, 88(6), 799–843. https://doi.org/10.3102/0034654318791584

Choi, J., Samadi, M. A., JaQuay, S., Park, S., & Nixon, N. (2025). *Read the Room or Lead the Room: Understanding Socio-Cognitive Dynamics in Human-AI Teaming* (arXiv:2510.09944). arXiv. https://doi.org/10.48550/arXiv.2510.09944

Erkan, A. (2019). Impact of Using Technology on Teacher-Student Communication/Interaction: Improve Students Learning. World Journal of Education, 9(4), 30–40.

Guha, A., Karumbaiah, S., Baeza, C., Castro, M., White, K., Roman, D. (2025a). Teacher Perceptions on AI Support for Bi/multilingual Learners: Superpowers and Constraints. *Artificial Intelligence in Education. AIED 2025. Lecture Notes in Computer Science, vol 15881*. https://doi.org/10.1007/978-3-031-98462-4_11

Guha, A., Kim, G., Karumbaiah, S., Bettney Heidt, E., Castro, M. (2025b) AI Support for Translanguaging in International School Classrooms. *International Conference for Computers in Education. ICCE 2025.* https://library.apsce.net/index.php/ICCE/article/view/5942

Holstein, K., McLaren, B. M., & Aleven, V. (2019). Co-Designing a Real-Time Classroom Orchestration Tool to Support Teacher–AI Complementarity. *Journal of Learning Analytics*, *6*(2), 27–52. https://doi.org/10.18608/jla.2019.62.3

Järvelä, S., Hadwin, A., Malmberg, J., & Miller, M. (2018). Contemporary Perspectives of Regulated Learning in Collaboration. In F. Fischer, C. E. Hmelo-Silver, S. R. Goldman, & P. Reimann (Eds.), *International Handbook of the Learning Sciences* (1st ed., pp. 127–136). Routledge. https://doi.org/10.4324/9781315617572-13

Jeong, H., & Hmelo-Silver, C. E. (2016). Seven Affordances of Computer-Supported Collaborative Learning: How to Support Collaborative Learning? How Can Technologies Help? *Educational Psychologist*, *51*(2), 247–265. https://doi.org/10.1080/00461520.2016.1158654

Lawrence, L., Garcia, D. I., Ochoa, K., Valdez, S., Begolli, K., Bustamante, A. S., & Ahn, J. (2023). *Factors That Impact the Implementation of a Game Based Curriculum and Adaptations Teachers Design to Address Them*. 918–921. https://doi.org/10.22318/icls2023.323293

Miller, M., & Hadwin, A. (2015). Scripting and awareness tools for regulating collaborative learning: Changing the landscape of support in CSCL. *Computers in Human Behavior*, *52*, 573–588. https://doi.org/10.1016/j.chb.2015.01.050

Smith, B. L., & MacGregor, J. T. (1992). What is collaborative learning.

van Leeuwen, A., Janssen, J., Erkens, G., & Brekelmans, M. (2015). Teacher regulation of cognitive activities during student collaboration: Effects of learning analytics. *Computers & Education*, *90*, 80–94. https://doi.org/10.1016/j.compedu.2015.09.006

Wei, X., Wang, L., Lee, L.-K., & Liu, R. (2025). The effects of generative AI on collaborative problem-solving and team creativity performance in digital story creation: An experimental study. *International Journal of Educational Technology in Higher Education, 22(1), 23.* https://doi.org/10.1186/s41239-025-00526-0

Wise, A.F., Schwarz, B.B. Visions of CSCL: eight provocations for the future of the field. *Intern. J. Comput.-Support. Collab. Learn* 12, 423–467 (2017). https://doi.org/10.1007/s11412-017-9267-5

Acknowledgements

The research reported in the article was made possible (in part) by a grant from the Spencer Foundation (#202500160). The views expressed are those of the authors and do not necessarily reflect the views of the Spencer Foundation. Support for this research was provided by the Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin-Madison with funding from the Wisconsin Alumni Research Foundation. We thank the Multilingual Learning Research Center at UW-Madison for supporting the recruitment process, Dr. Ananya Ganesh and Kaycie Barron for their review on our thematic analysis, and Dr. Devansh Saxena, Dr. Esther Bettney Heidt, and Dr. Mariana Castro for their valuable feedback on our study design and protocol. We thank the CSCL reviewers for their feedback and our teacher participants for sharing their valuable time and insights with us.

International Society of the Learning Sciences

Integrating Augmented Reality and Product Design in Makerspaces

Dejan Trencevski, Jina Kang, Vishal Sachdev, Aric Rindfleisch
dtren2@illinois.edu, jinakang@illinois.edu, vishal@illinois.edu, aric@illinois.edu
University of Illinois at Urbana Champaign

Abstract: This study aims to explore the affordances of augmented reality technologies (AR) to facilitate students' product design processes in a business course. We analyze the affordances of AR in the context of a product design activity in an undergraduate classroom. The course is held in a Makerspace lab at a university campus, and it requires students to conceptualize, design, prototype, manufacture, and market new products by using 3D computer-aided design (CAD) software and 3D printing. Our study illuminates that although CAD allows students to visualize and manipulate digitally manufactured 3D objects on a 2D computer monitor surface, it does not allow them to visualize and manipulate those 3D objects in 3D space. However, AR does afford visualization and manipulation of digitally manufactured 3D objects in 3D space which is why we investigated how do interactions with 3D models in AR impact product design processes for students.

Introduction

Developing students' ability to visualize and manipulate three-dimensional (3D) objects is essential for design and engineering education, yet traditional computer-aided design (CAD) tools constrain learners to 2D screens. These constraints can limit students' spatial reasoning and make it difficult for teams to achieve shared understanding of designs when collaborating. Recent advances in augmented reality (AR) provide new possibilities for supporting these processes by enabling learners to view and interact with 3D models in real physical space (Garzon & Acevedo, 2019). Due to its immersive visualization and shared spatial reference capabilities, AR can make abstract design concepts more tangible and promote visualization.

Prior research has documented the pedagogical potential of AR for enhancing spatial cognition and engagement in science learning (Cheng & Tsai, 2013). However, studies examining AR-supported product design, particularly within Makerspace learning environments, remain limited. Existing work has primarily emphasized technical implementation or learner motivation, with less attention on how students experience and interpret AR's affordances for understanding and collaboration during authentic product design tasks.

To address this gap, we conducted an exploratory study that investigates how college students engage with 3D modeling tasks supported by AR technology. This study took place in a Makerspace course that integrates CAD and 3D printing. After completing a CAD-based design task, students engaged in a similar activity using AR headsets to interact with their 3D models in an immersive manner. The researchers then examined how students described their individual and team-based experiences with AR and what affordances they perceived as supporting spatial understanding and collaboration during product design. Grounded in spatial cognition and social constructivist perspectives, we view AR as a tool that can make spatial relationships perceptually accessible and provide a shared visual field for the co-construction of ideas. Our goal was not to compare learning outcomes between CAD and AR, but to explore students' perceptions of how AR-mediated interaction might enhance visualization and teamwork in an actual design-based learning context. Specifically, we sought to answer two key research questions:

RQ1: How do students describe their experiences interacting with 3D models using AR during individual and team-based product design activities?

RQ2: What affordances of AR-supported interaction do students perceive as supporting their spatial understanding and team collaboration during product design?

Background and theoretical framework

Spatial cognition

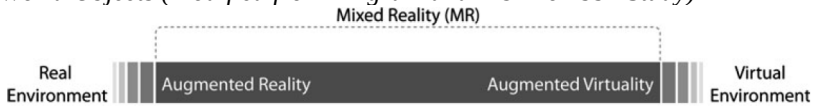
AR is the technology that overlays virtual elements in the real world (Akcayir & Akcayir, 2017; Garzon & Acevedo, 2019). This technology has been shown to have potential for pedagogical applications for over a decade. In a literature review of empirical research on affordances of augmented reality in science learning, Cheng & Tsai (2013) suggest that AR affords students' spatial ability, particularly by supporting spatial cognition, which refers to how individuals perceive, represent, and reason about spatial properties like size, shape, location or direction

 © ISLS

of objects or events in the world. AR affordances are features or properties of elements that present cues on what can be done with them and provide input on how students can interact with them (Shin, 2021). In this sense, AR can reduce reliance on internal mental visualization by externalizing spatial representations, allowing learners to directly perceive and manipulate 3D forms. The mix of physical and virtual elements in the coexisting physical and virtual learning environments create a new type of reality for students, a mixed reality (MR). The MR and AR definitions regularly adopted by empirical research are described as variations of Virtual Reality (VR) publicized by Azuma as shown in Figure 1 (Azuma, 1997; Cheng & Tsai, 2013). Situated along the virtuality continuum (Milgram & Kishino, 1994), these environments enable iterative interactions between perception and action, which are central to developing spatial understanding in design contexts.

Figure 1

Virtuality Continuum Defining the Possibilities of the Mixture of Real-World and Virtual-World Objects (Modified from Milgram and Kishino 1994 Study)



Social constructivism

Vygotsky's (1978) theory of social constructivist learning emphasized the social and collaborative nature of learning. A learning environment that promotes groups collaborating and constructing knowledge for one another through social interaction would thus embody the philosophy of social constructivist learning. Cheng and Tsai (2013) specifically examined the theory's principles for development of educational activities among small groups and connected the pedagogical potential of the unique affordances provided by AR technologies. Throughout the literature review, Cheng and Tsai (2013) highlight the social interaction inherent in an AR activity that can provide opportunities for learning collectively as a group. AR allows students to interact both with their peers and the virtual environment, constructing a new understanding of the underlying data as the virtual objects respond to user input (Lan, C., et al, 2013; Taber, K.S., 2018). Through shared interaction with these representations, learners can coordinate attention, negotiate meaning, and build joint understanding in real time. In this way, AR environments function as mediating space where collaborative sensemaking is anchored in shared 3D artifacts.

Framing the learning problem and design approach

To study the potential pedagogical application of AR technology for product design in Makerspaces, our research team analyzed how students' product design processes are impacted when they are interacting with 3D objects in these immersive environments as individual learners and when they are socially interacting with other members of the team. The multimodal AR-aided experimental intervention that we designed utilizes hardware such as AR headsets, laptops, and desktops to facilitate students' product design processes while they are designing physical objects which will be 3D printed.

Teaching students (our future workforce) how to manufacture physical products can be a challenging experience for instructors. Often instructors don't have the desired manufacturing equipment required to teach the students the necessary skills they need to learn about how to manufacture physical products. Tools for milling, drilling, grinding, and finishing are vital artifacts that teachers use. However, manufacturing equipment can be costly, timely to use or set up, difficult to use or set up, and sometimes inadequate to achieve learning goals due to its physical nature. Therefore, it is important that instructors take advantage of all available resources to them. One available resource that helps instructors teach students how to manufacture physical products is by using 3D printers to enable quick prototyping. However, creating models for 3D printing requires knowledge of CAD tools, which have become easier to use over time but still require you to work in 2D.

Due to the limitations of working in 2D while trying to design a 3D model, students struggle with spatial reasoning in CAD because it requires them to imagine and approximate how their 3D model would look like in the real world rather than just seeing the actual 3D model in the real world. Designing in 2D becomes increasingly more difficult when working in teams due to lack of shared 3D understanding amongst team members. This creates a dual challenge: students must internally translate between 2D representations and 3D outcomes while also attempting to align these interpretations across team members. From a spatial cognition perspective, this increases cognitive load by requiring internal visualization rather than direct interaction with 3D forms. From a social constructivist perspective, it constrains collaboration by limiting access to shared, manipulable representations that support joint attention and coordinated action. But AR does afford interactions with 3D

objects in 3D space in the context of Makerspaces which can afford students the opportunity to measure, identify, understand, and practice seeing general principles at work while using their spatial abilities in the real world, thus creating a joined understanding of the problems that students are trying to solve (NCTM, 2022). By externalizing design artifacts into the physical environment, AR enables learners to directly perceive and manipulate 3D models, supporting both individual spatial reasoning and shared interpretation within teams.

Extended reality (XR) technology, such as augmented reality (AR) and virtual reality (VR), has been listed as critical and emerging technology by the National Science Foundation (NSF) and it has been highlighted as one of the most innovative pedagogies in recent empirical reports (National Science and Technology Council, 2022; Kukulska-Hulme et al., 2021). Social interaction among students who are immersed in environments with interactive 3D objects in 3D space could allow for AR-aided learning to occur. Uniting the affordances of Makerspaces 3D printing technology with AR technology could not only enhance the learning experience, but it also has potential for transformational pedagogical applications in science learning (Cheng & Tsai, 2013; McQuiggan, 2015). To explore this potential, we designed an intervention to test the effects of 3D modeling with CAD in AR-facilitated product design activities situated in a Makerspace lab with easy access to 3D printers. The intervention was designed to allow students to view and manipulate their CAD-created 3D model in a 3D space with an AR headset as a form of iteration before they physically print them on a 3D printer and spend physical materials. This study therefore examines how AR-mediated interaction with 3D models shapes students' spatial reasoning and collaborative product design processes in this context.

Approach

Utilizing 3D printers as a resource for teaching students how to design, manufacture, and market physical objects could serve as a substitute for costly physical manufacturing equipment. Makerspaces can serve as catalyzers for innovation where students turn ideas to objects using 3D computer-aided design (CAD) software. Traditional 3D modeling with CAD only allows for visualizing 3D objects on a 2D plane which means that the users' spatial abilities are distorted since they can't visualize the third dimension, the dimension of depth (or height). Further, these technologies limit the viewpoint to a small 2D computer screen rather than a large 3D space we typically embody while we communicate and interact in person. That means that our human sense of vision is limited by the capabilities of technology. However, those 3D CAD objects could be viewed and manipulated in AR. AR technology will allow students to visualize 3D objects in 3D space.

Methods

Participants and context

The participants in the study's intervention were from an undergraduate business course that is a hands-on course in which interdisciplinary teams of business, design, and engineering students conceptualize, design, prototype, manufacture, and market new products. To create these products, they use 3D software and hardware. By enrolling in this course, students develop collaboration, design, manufacturing, and marketing skills. This course is held in a Makerspace at a university campus, and it typically has an enrollment of 21 students that are split into 7 teams of 3 members each. Of the 21 students that were enrolled in the course, 16 students participated in our study.

For this study we designed an experimental intervention that incorporated AR into the existing Makerspace learning environment while students are learning how to do 3D modeling with CAD. The intervention was designed to allow students to view and manipulate their CAD-created 3D model in a 3D space with a Microsoft HoloLens2 AR headset as a form of iteration before they physically print them on a 3D printer and spend physical materials. The HoloLens2 headset is a lightweight mixed reality device that users would wear on their head like a pair of glasses. This device allows users to not only see the world around them but also see digital augments in their field of view. Students wore the HoloLens2 AR headset and interacted with the virtual space with simple manipulations such as buttons on the headset or with hand gestures. Each team had one AR headset that was set up to cast a live view of the AR headset to a nearby computer monitor. That meant that, regardless of who was wearing the AR headset, every member of the team had a live view of the same 3D object in that same 3D space. The intervention enabled students to interact with their 3D model in 3D space by moving it, rotating it, and scaling it, giving them the opportunity to construct products that had been previously unrealized by traditional tools. Manipulating 3D objects in 3D space afforded students the opportunity to measure, identify, understand, and practice seeing general principles at work while using their spatial abilities and socially interacting in the real world (NCTM, 2000; Topraklikoglu & Ozturk, 2021).

Procedure and data sources

 © ISLS

Data was collected from students who filled out paper questionnaires forms after day two of the study. Participants answered two open-ended questions to describe how interactions with 3D models in AR impacted their product design process as an individual and as a team:

"How did interactions with 3D models in AR impact your product design process as an individual?"

"How did interactions with 3D models in AR impact your product design process as a team?"

The intervention took place during the students' regularly scheduled class instruction. Our intention was to make minimal disruption and as small changes as possible to the natural learning environment. It is standard practice in this course for students to complete 3D modeling with CAD activities and questionnaires about their education. Thus, integrating this intervention into the classroom activities aligned well with the regular curriculum because it comprised of 3D modeling with CAD + AR and a questionnaire.

The course curriculum guides the students from their ideas about products to real physical objects that they 3D print and market. As shown in the semester syllabus timeline in Figure 2, students spend week 5 through week 7 on conceptualizing their products. During this three-week period, they draw and sketch their products on paper as a precursor to designing their products in CAD the following two weeks. Figure 3 provides examples of their concept drawings.

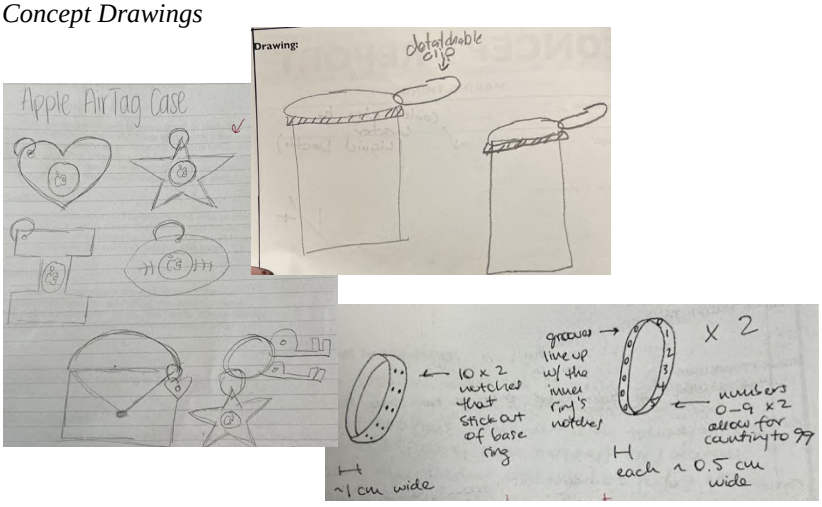
Figure 2

Semester Timeline

DATE	TOPIC	DELIVERABLES
WEEK 1	Course Overview	Capability Assessment
WEEK 2	Target Audience & 3D Printing	Target Analysis
WEEK 3	Making Opportunities & 3D Slicing	Opportunity Analysis
WEEK 4	Ideas of What to Make & 3D Design	Idea Inventory
WEEK 5	Initial Concept	Concept Audit
WEEK 6	Concept Testing	Concept Test Results
WEEK 7	Concept Finalization	Concept Report
WEEK 8	Designing your Product	Design File
WEEK 9	Design Assesment	Design Audit
WEEK 10	Prototyping your Product	Prototype Builds
WEEK 11	Prototype Testing	Prototype Test Report
WEEK 12	Refining your Product	Updated Design & Build
WEEK 13	Finalizing your Product	Final Design & Build
WEEK 14	Marketing your Product	M & M Plan
WEEK 15	Project Presentations	Present/Attend

Figure 3

Concept Drawings



During day one of the study (week 8 in the semester), the participants completed a 3D modeling activity with CAD. Students were working in groups of three and their task was to use CAD and develop a 3D model of their product that they were going to 3D print. A post-activity questionnaire was administered to evaluate the students' attitudes and perceptions towards 3D modeling with CAD during product design (see Figure 4).

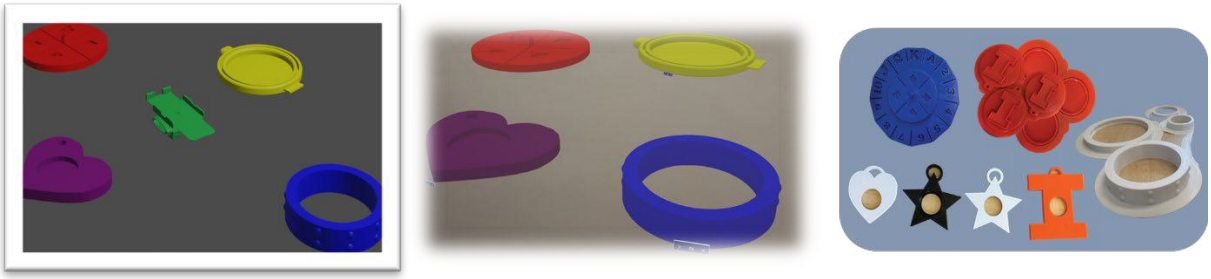
 © ISLS

During day two of the study (week 9 in the semester), the same 3D modeling with CAD activity and procedure as day one occurred, with the addition of AR headsets for students to visualize their 3D models. A post-activity questionnaire was administered to evaluate the students' attitudes and perceptions towards 3D modeling with CAD + AR during product design (see Figure 4).

This allowed the classroom activities to flow as smoothly as possible and the teams to advance their designs toward creative 3D printed models illustrated in Figure 4 that included Apple AirTag case holders, fidget rings, board game pieces, and soda can lids.

Figure 4

3D Models in CAD, 3D Models in AR, and Printed 3D Models



Analysis

We conducted an inductive thematic analysis (Braun & Clarke, 2006) of students' open-ended responses to explore how they experienced and described their interactions with 3D models in AR. While our coding process was data-driven and exploratory, we drew on frameworks from spatial cognition (Hegarty & Waller, 2004) and social constructivism (Vygotsky, 1978; Dillenbourg, 1999) to inform our interpretation of emerging patterns. These theories guided how we grouped and named themes, particularly in relation to how students described their spatial reasoning and collaborative practices during the design process.

Two researchers independently coded the questionnaire responses, identifying patterns in how students described their individual and team experiences with AR. Initial codes were compared and refined through discussion, resulting in three primary themes: spatial visualization, collaborative awareness, and immersive interaction. Each theme was defined based on recurring patterns in student language and the types of affordances students perceived as supporting their design work.

Results

Spatial visualization

Students described AR as enabling them to perceive their designs as tangible objects in physical space, rather than as abstract representations on screens. This theme captures how AR supported students' ability to assess scale, proportion, and spatial relationships without requiring mental rotation or imagination.

Many students described how seeing their 3D models in AR supported their ability to visualize the product's shape, scale, and structure in physical space. Unlike working in CAD alone, which required mental rotation or imagining proportions, students reported that AR helped them "see" the model as if it were already built:

> "We were all able to *visualize* the product before printing it and seeing where we can develop on our design and next actions to be taken in regards to our product and design."

> "*Seeing* our model using the AR technology helped us get a feel of what our product would *look* like in real life and helped us understand what areas we should change."

These comments suggest that AR helped bridge the gap between digital modeling and embodied spatial reasoning. Rather than relying on mental transformation of 2D views into 3D understanding, students could directly perceive spatial relationships, making design evaluation more concrete and action-oriented. The shift from "imagining" to "seeing" represents a reduction in cognitive demand—students no longer had to mentally construct the third dimension.

 © ISLS

Prior research in spatial cognition highlights the importance of external visualizations in supporting mental manipulation and design planning (Hegarty & Waller, 2004; Johnson-Glenberg et al., 2016). Our findings suggest that AR may offer similar benefits by providing situated, scale-aware representations that reduce the cognitive demand of imagining 3D transformations. By making spatial properties perceptually accessible rather than inferential, AR may support more direct spatial reasoning during design iteration.

Collaborative awareness

This theme captures how AR enabled shared perception of design objects, allowing team members to reference the same spatial features simultaneously and coordinate their attention during design discussions. Students described AR as supporting both individual spatial understanding and collective decision-making. Eight students explained how AR enabled team collaboration and communication during the design process:

> "As a team AR facilitated *collaborative* decision-making by providing a shared immersive experience where we could all examine the 3D model from different angles. It improved communication, as we could point out issues directly on the model, making discussions more efficient."

> "The interactions improved visualization and realism. They enhanced *collaboration*. They helped with design accuracy and functional products."

The emphasis on "shared immersive experience" and the ability to "point out issues directly on the model" suggests that AR provided a common spatial reference field that grounded collaborative discourse. Rather than each team member maintaining their own mental model and attempting to coordinate through verbal description alone, AR created a shared perceptual space where spatial features could be directly indicated and jointly attended to.

Considering that 50% of our participants' responses supported decades of empirical research, social interaction itself enables learning and has a positive effect especially for young learners. Along the same lines, positive effects of collaborative learning where students work together toward a common learning goal instead of a solo learning environment have been known for a very long time in the Learning Sciences (Hersh-Pasek K., et al, 2015; Johnson, D.W., et al, 1981). Thus, it was a pleasure to see that this type of collaboration is also supported in AR environments where students reported that they enjoyed the shared immersive experiences and also be aware of the physical outside environment.

Immersive interaction

This theme captures how students experienced AR's interactive capabilities – particularly the ability to manipulate, navigate around, and physically engage with virtual objects in 3D space. Students described these interactions as more intuitive and effective than traditional CAD interfaces, emphasizing how direct manipulation in immersive space supported their design evaluation. Eight students highlighted the importance of interacting with 3D objects in 3D space through the AR interface:

> "*Interacting* with 3D models in AR helped me visualize the product in a *real world* environment, allowing me to assess its scale, properties, and functionalities more effectively than with traditional sketches or digital models."

> "As an individual, the *interactions* with the 3D model allowed me to navigate and "*feel*" our prototype in an easier and more intuitive manner than the 3D viewing possible when using a model in Fusion 360."

These responses reveal a distinction between passive viewing and active interaction. The emphasis on being able to "navigate" and "feel" the prototype suggests that AR's value extends beyond visualization alone—it includes embodied interaction with virtual objects. The comparison to Fusion 360 (a CAD program) is particularly telling: while CAD allows rotation and zoom through mouse and keyboard controls, AR enables manipulation through natural gestures and physical movement around the object. This shift from interface-mediated control to direct manipulation may make the design evaluation process "easier and more intuitive."

The student's use of "feel" in quotation marks is noteworthy. While AR does not provide haptic feedback, students described a sense of tangibility or material presence when interacting with virtual objects in physical

space. This suggests that spatial co-location and gestural interaction can create a psychological sense of physical engagement even without touch, supporting more concrete design reasoning.

Research on embodied cognition suggests that physical interaction with objects supports deeper conceptual understanding than passive observation (Johnson-Glenberg et al., 2016). By enabling gestural manipulation and navigation in 3D space, AR may activate sensorimotor schemas that support spatial reasoning. The ability to walk around objects, rotate them with hand gestures, and view them from multiple perspectives mirrors how we naturally explore physical objects in the world. This alignment between AR interaction and everyday spatial behavior may explain why students found AR "more intuitive" than screen-based CAD tools that require learning arbitrary interface conventions.

Furthermore, the immersive quality of AR—where virtual objects appear embedded in the physical environment—may reduce the psychological distance between design representation and fabricated product. This could support what cognitive scientists call *near transfer*, where learning in one context (AR visualization) transfers more readily to another context (physical product evaluation) because the two contexts share perceptual and spatial features.

Discussion and conclusion

This study explored how undergraduate students experienced AR technology during product design activities in a Makerspace course. We analyzed students' descriptions of their experiences and identified three primary themes: spatial visualization, collaborative awareness, and immersive interaction. Our findings contribute to understanding how AR can support both spatial reasoning and collaborative work in design-based learning environments.

AR and spatial cognition

Our findings show that AR reduced the cognitive demands students faced when working with 3D designs. Students repeatedly described AR as enabling them to "see" their products "in real life" rather than having to imagine what their 2D CAD designs would look like in 3D space. This matters because traditional CAD tools require students to mentally rotate, scale, and imagine depth—cognitive operations that can be difficult and error-prone. AR externalizes these spatial relationships by making them directly visible in physical space.

This finding extends prior work on AR in science education (Cheng & Tsai, 2013; Garzón & Acevedo, 2019) by showing a specific way AR supports learning: students perceive reductions in cognitive demands with mental transformations. Instead of viewing AR simply as a motivational tool, our results suggest AR's core value is making spatial properties directly perceivable. This is particularly important for design education, where spatial reasoning is central and where learners often struggle with the transition from 2D sketching to 3D fabrication.

Students emphasized being able to assess "scale" and "proportions" in physical space. This highlights how spatial understanding is grounded in our bodily experience of navigating and interacting with objects (Johnson-Glenberg et al., 2016). By placing virtual objects in physical space, AR taps into students' natural abilities for perceiving depth, distance, and scale more effectively than screen-based tools.

Supporting collaborative design

Our second major finding was that students valued AR for supporting teamwork. Eight students described AR as facilitating collaborative decision-making through "shared immersive experiences" where team members could "examine the 3D model from different angles" and "point out issues directly on the model." This shows how AR functions as a mediating tool that supports joint attention and coordinated activity (Vygotsky, 1978; Roschelle, 1992).

What stands out is how AR creates a shared visual space that grounds team discussions. In traditional CAD workflows, each team member views the model on their own screen, and coordination requires verbal description or screen-sharing. AR allows team members to view and manipulate a single virtual object together in shared physical space. This shared visual reference supports the development of common understanding through joint activity (Roschelle, 1992).

This finding adds to existing research on AR and collaboration (Lan et al., 2013) by identifying a specific collaborative affordance: the creation of shared spatial reference. AR enables pointing, joint attention, and simultaneous manipulation—all central to effective design collaboration (Dillenbourg, 1999). This suggests AR may be particularly valuable in learning contexts that emphasize collaborative problem-solving around spatial or physical artifacts.

Design implications for learning

Our findings suggest several practical applications for educators integrating AR into Makerspaces.

 © ISLS

AR as an intermediate design tool

Rather than replacing CAD or 3D printing, AR works best as an intermediate visualization and evaluation tool. Students in our study used AR after creating CAD models but before 3D printing, allowing them to iterate on designs without wasting physical materials. This suggests a design workflow where AR serves as a *spatial preview* stage that reduces costly fabrication errors.

Supporting collaborative design reviews

The collaborative benefits of AR suggest its value for design critique and peer feedback activities. Educators can structure design review sessions where teams use AR to present their work, with peers providing feedback while viewing the AR model together. This formalizes the collaborative benefits students described and supports peer learning.

Scaffolding spatial reasoning

For students with less developed spatial abilities, AR provides essential support during the transition from conceptual sketches to fabricated objects. Instructors might provide additional AR visualization opportunities for students who struggle with mental rotation or spatial transformation tasks.

Design for shared attention

When implementing AR in collaborative contexts, prioritize features that support joint attention—for example, multi-user AR experiences where team members see the same virtual object simultaneously, or annotation tools that allow collaborative marking and discussion.

Limitations and future research

Our study has several limitations. First, our sample was small (sixteen students) from a single course, limiting generalizability. Second, we relied on self-report data through open-ended questionnaires. While this provided insight into students' experiences, we cannot assess actual learning outcomes or compare AR-supported learning with traditional instruction. Students may have valued AR for novelty rather than learning benefits. Third, we did not include a comparison condition or measure spatial reasoning abilities directly. We cannot make claims about AR's causal effects on learning. Fourth, our thematic analysis was interpretive, and other researchers might identify different themes.

Future research should address these limitations in several directions. We plan to iterate on this intervention across multiple course implementations using design-based research methods (Brown, 1992), gathering both qualitative and quantitative data on learning outcomes. This includes pre-post assessments of spatial ability, comparative analysis of design artifacts, and video analysis of collaborative design sessions.

Future studies should include comparison conditions (CAD only vs. CAD + AR) and measure learning outcomes such as design quality, spatial reasoning test scores, and team communication effectiveness. Longer-term implementation would show whether AR's benefits continue as novelty fades and whether students internalize spatial reasoning strategies. Research should also investigate whether AR supports equitable learning across diverse student populations and examines differential effects by prior spatial ability or domain expertise.

Conclusion

This study explores how AR can support both spatial reasoning and collaborative work in design-based learning contexts. Our findings suggest AR's value lies in transforming the nature of spatial reasoning and collaborative work during design activities. By making spatial relationships directly perceivable and creating shared visual fields for collaboration, AR addresses two challenges in design education: the cognitive demands of working across 2D and 3D representations, and the coordination challenges of collaborative design work.

As AR hardware becomes more affordable and accessible, and as educators prepare students for futures where mixed reality technologies are commonplace, understanding AR's pedagogical affordances becomes increasingly important. Our analysis of this AR-integrated product design activity offers a model for incorporating AR into Makerspace curricula, grounded in spatial cognition and social constructivist theories. We hope this model will guide other educators and researchers in designing and studying AR-supported learning experiences.

The future of work increasingly involves designing, prototyping, and collaborating in mixed reality environments. Similar to how personal computers, laptops, and smartphones evolved from emerging technologies to essential tools, we anticipate that today's immersive head-mounted displays will become essential in the future. By investigating how AR supports learning in contexts like Makerspaces, we contribute to preparing students with the spatial reasoning and collaborative capabilities essential for this emerging landscape.

References

Akçayır, M., & Akçayır, G. (2017). Advantages and challenges associated with augmented reality for education: A systematic review of the literature. Educational Research Review, 20, 1–11. https://doi.org/10.1016/j.edurev.2016.11.002

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Brown, A. L. (1992). Design Experiments: Theoretical and Methodological Challenges in Creating Complex Interventions in Classroom Settings. Journal of the Learning Sciences, 2(2), 141–178. https://doi.org/10.1207/s15327809jls0202_2

Cheng, K.-H., & Tsai, C.-C. (2013). Affordances of Augmented Reality in Science Learning: Suggestions for Future Research. Journal of Science Education and Technology, 22(4), 449–462. https://doi.org/10.1007/s10956-012-9405-9

Critical and Emerging Technologies List 2024 Update | OSTP. (2024, February 12). The White House. https://www.whitehouse.gov/ostp/news-updates/2024/02/12/critical-and-emerging-technologies-list-2024-update/

Dillenbourg, P. (1999). What do you mean by collaborative learning?. *Collaborative-learning: Cognitive and computational approaches.*, 1-19.

Garzón, J., & Acevedo, J. (2019). Meta-analysis of the impact of Augmented Reality on students' learning gains. Educational Research Review, 27, 244–260. https://doi.org/10.1016/j.edurev.2019.04.001

Hegarty, M., & Waller, D. (2004). A dissociation between mental rotation and perspective-taking spatial abilities. Intelligence, 32(2), 175–191. https://doi.org/10.1016/j.intell.2003.12.001

Hirsh-Pasek, K., Zosh, J. M., Golinkoff, R. M., Gray, J. H., Robb, M. B., & Kaufman, J. (2015). Putting Education in "Educational" Apps: Lessons From the Science of Learning. Psychological Science in the Public Interest, 16(1), 3–34. https://doi.org/10.1177/1529100615569721

Johnson-Glenberg, M. C., Megowan-Romanowicz, C., Birchfield, D. A., & Savio-Ramos, C. (2016). Effects of embodied learning and digital platform on the retention of physics content: Centripetal force. Frontiers in Psychology, 7, 1819. https://doi.org/10.3389/fpsyg.2016.01819

Johnson, D. W., & Johnson, R. T. (1981). Effects of cooperative and individualistic learning experiences on interethnic interaction. *Journal of Educational Psychology, 73*(3), 444–449. https://doi.org/10.1037/0022-0663.73.3.444

Johnson, L., Adams, S., & Cummins, M. (2012). NMC Horizon Report: 2012 K-12 Edition. New Media Consortium. https://eric.ed.gov/?id=ED593595

Kukulska-Hulme, A., Bossu, C., Coughlan, T., Ferguson, R., FitzGerald, E., Gaved, M., Herodotou, C., Rienties, B., Sargent, J., Scanlon, E., Tang, J., Wang, Q., Whitelock, D., & Zhang, S. (2021). Innovating Pedagogy 2021: Open University Innovation Report 9 [Other]. The Open University. https://iet.open.ac.uk/innovating-pedagogy

Lan, C. H., Chao, S., & Chao, K. H. (2013). Mobile Augmented Reality in Supporting Peer Assessment: An Implementation in a Fundamental Design Course. *International Association for Development of the Information Society*.

McQuiggan, S. (with Kosturko, L., McQuiggan, J., & Sabourin, J.). (2015). Mobile learning: A handbook for developers, educators, and learners. John Wiley & Sons.

Milgram, P., & Kishino, F. (1994). A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems, 77*(12), 1321-1329.

National Council of Teachers of Mathematics (2000). Principles and evaluation standards for school mathematics. Virginia: Reston.

National Science and Technology Council (2022). Critical and Emerging Technologies List Update.

Roschelle, J., & Clancey, W. J. (1992). Learning as Social and Neural. *Educational Psychologist, 27*(4), 435–453. https://doi.org/10.1207/s15326985ep2704_3

Shin, D. (2022). Does augmented reality augment user affordance? The effect of technological characteristics on game behaviour. Behaviour & Information Technology, 41(11), 2373–2389. https://doi.org/10.1080/0144929X.2021.1928286

Topraklikoglu, K., & Ozturk, G. (2021). Teaching Geometry through Augmented Reality Activities Using Three-Dimensional Modelling. International Online Journal of Educational Sciences. https://doi.org/10.15345/iojes.2021.05.003

Vygotsky, L.: Mind and Society by Lev Vygotsky. (1978).

Short Papers

 © ISLS

 © ISLS

International Society of the Learning Sciences

Unpacking Flexibility in Teacher Professional Learning on WhatsApp Messaging Network Platform

Iris Saban, University of Haifa, irisgreenhouse@gmail.com
Linor L. Hadar, University of Haifa, lhadar@edu.haifa.ac.il

Abstract: This study examines how flexibility manifests as a lived practice within a voluntary WhatsApp-based teachers' professional community. While flexibility is a recognized affordance of online learning, its role in teacher professional networks remains undertheorized. Drawing on 2,307 posts across 349 conversations among 146 teachers, we conceptualized and measured flexibility through three dimensions: temporal, concerning engagement during discretionary time; sequential, involving the recurrence of conversations; and structural, which maps network connections across discussions. The analysis reveals profound temporal flexibility, with nearly two-thirds of posts occurring outside school hours and over a 40% during holidays and summer break. Although recurring conversations constituted 11% of discussions, they generated 25.4% of posts and were longer with statistically significant differences. These recurring discussions formed the network's core, effectively integrating teachers and sustaining dialogue. These findings demonstrate that flexibility is not a static design feature but a dynamic, socially constructed practice fundamental to professional growth.

Introduction

As teacher professional development migrates to online spaces, a critical need exists to understand the mechanisms that sustain learning in these environments. While teacher communities are celebrated for fostering collaboration (Hadar & Brody, 2017; Vangrieken et al., 2017), the integration of popular platforms like WhatsApp introduces under-examined dimensions of professional learning. This paper, part of a larger project, presents a focused investigation into one aspect: the role of flexibility in time, pace, and format. Building on Collis and Moonen's (2012) framework, this study adapts two of their dimensions, temporal and sequential, to the social network context and extends the model by introducing a structural dimension that captures the connective nature of conversations across the network. This initial focus allows for a deep analysis, while future stages may incorporate additional dimensions tailored to the community's dynamics. Drawing on theories of learning in communities (Lave & Wenger, 1991; Lantz-Andersson et al., 2018) and the affordances of online learning (Lay et al., 2020) and situating this study within a well-established literature on online teacher professional development and teacher communities (Vangrieken et al., 2017; M Macià & García, 2016), we analyze how teachers leverage WhatsApp's inherent flexibility. We argue that flexibility is not merely a logistical feature of online learning, but a central element that shapes participation and sustains teacher learning in the community.

Theoretical background

Teacher learning in professional communities

The pursuit of educational quality improvement depends on recruiting effective teachers and providing them with continuous, meaningful professional development (Barber & Mourshed, 2007). Teacher quality, however, is not a static attribute acquired in formal preparation; it is a dynamic capacity cultivated through continuous learning, a process best supported by collaborative practices (Author 2 & colleague, 2010). Sustained collaboration is essential for professional growth, reshaping both teaching methods and teachers' core conceptions of their practice (Shulman, 2005; Mourshed et al., 2011). Therefore, investigating effective methods for developing their knowledge is critical for strengthening educational systems.

This perspective resonates with Lave and Wenger's (1991) theory of communities of practice, which reframes learning as a social process situated within authentic interactions, where participation and mutual support are central. The core characteristics of teacher communities, including shared vision, collective responsibility, reflective dialogue, and trust, underscore the importance of non-hierarchical and sustained collaboration (Bolam et al., 2005; Vescio et al., 2008; Scherz et al., 2021). When supported by flexibility, the very conditions that enable and sustain teacher learning are amplified, allowing these collaborative features to manifest more dynamically.

 © ISLS

Online learning and virtual learning communities

The digital landscape offers significant opportunities for teacher development, including access to resources, flexibility, autonomy in regulating learning, and rich potential for interaction (Lay et al., 2020; Bragg et al., 2021). Yet, research consistently highlights the difficulty of cultivating authentic collaboration in online environments, a challenge contributing to high attrition rates in online programs (Bragg et al., 2021). Virtual learning communities can address this gap by offering spaces that are both structured and flexible. In these spaces, professional dialogue, trust, and mutual support are actively cultivated. Their unique advantage lies in fostering collaborative interactions while leveraging the distinct benefits of online learning, notably the affordance of flexibility.

Flexibility in online professional learning

Within online learning, flexibility is a defining affordance, commonly described as the provision of self-directed learning opportunities, enabling teachers to select content, regulate pace, and adapt the depth of engagement (Veletsianos & Houlden, 2019; Bragg et al., 2021). Flexible learning encompasses the ability to learn at any time, location, and to integrate different modalities (Loon, 2022; Li & Wong, 2018). This allows teachers to weave professional learning into their daily routines (Ulanday et al., 2021).

However, established frameworks for flexibility were mostly developed for formal, institution-based systems and are less applicable to dynamic social platforms (Collis & Moonen, 2012). Collis and Moonen conceptualized flexibility across five dimensions: (1) *time*, referring to when learning takes place and the pace at which learners progress; (2) *location*, where learning occurs, whether at home, school, or workplace; (3) content, what subject matter is studied and to what depth; (4) *instructional approach*, how learning activities are designed and delivered; and (5) *delivery and assessment*, how learning outcomes are evaluated and recognized. While time and location transfer readily to social networks, the other dimensions require significant adaptation. In WhatsApp-based communities, for example, content emerges organically from participants rather than from a predefined syllabus, instructional methods are socially negotiated rather than instructor-driven, and formal assessment is replaced by social recognition and continued participation. Applying this framework to informal, networked contexts highlights both its enduring relevance and the need for conceptual extensions. This theoretical gap is mirrored in empirical work. Studies measuring flexibility (e.g., Soffer et al., 2019) have largely been situated in higher education, conceptualizing it as a technical attribute of learning management systems rather than as a socially constructed practice. Most studies frame flexibility in relation to efficiency, access, or user satisfaction (Huang et al., 2024), leaving its role in the learning process itself underexplored. This underscores the need to investigate how flexibility is enacted in practice. The present study addresses this gap by examining a WhatsApp-based professional community as a site where flexibility is continuously constructed. The analysis is part of a larger ongoing project.

Research questions

The primary question guiding this study is: How is flexibility in teacher learning enacted as a lived practice within a WhatsApp-based professional community? Specifically, we ask three research questions: RQ1: At what times do teachers engage in the community, and how does this reflect their use of flexibility (temporal dimension)? RQ2: How does the recurrence of conversations manifest over time, and what is its significance for learning (sequential dimension)? RQ3: How are network connections across conversations shaped by teachers participating in multiple discussions, and what does this indicate about patterns of continuity in the community (structural dimension)?

Methods

Context and participants

The study examines a voluntary virtual professional learning community for middle school teachers of Jewish-Israeli Culture Studies, a newly introduced mandatory subject for which most teachers had no prior training. Established in 2017 by the Center for Educational Technology (CET), the community operates on WhatsApp and is facilitated by two facilitators, who combine structured routines, synchronous discussions, and ongoing peer interaction. The analysis covered one school year (2021-2022), during which 149 teachers in the community actively contributed at least one post.

Data sources and units of analysis

 © ISLS

We analyzed 2,307 posts (containing an identifier, timestamp, teacher identity, and message content) organized into 349 conversations. Drawing on Hymes's (1972) ethnography of communication, we treat conversations as bounded speech events defined by continuity in time, content, and participation. Building on discourse research highlighting the role of sequences in shaping meaning (Lefstein et al., 2015), we defined a conversation by its thematic continuity, classifying messages as part of the same conversation if teachers returned to a previous point, after extended pauses, sometimes with new participants. This extends the definition of conversation boundaries to include thematic continuity in addition to temporal and sequential criteria.

Modes of analysis

To capture flexibility, we adapted existing frameworks (e.g., Collis & Moonen, 2012) to this social network context, examining three dimensions: (1) Temporal flexibility: This dimension captures the ability to participate at any given time. We employed descriptive statistics to analyze post timestamps, tracing teachers' activity beyond school hours, including different days, time segments, holidays, and summer break. (2) Sequential flexibility: This refers to the capacity to sustain dialogue over time. It was measured by identifying conversation recurrence and by comparing post averages in recurring versus non-recurring discussions. Due to non-normal data distribution, a Mann–Whitney U test was applied to assess statistical significance. (3) Structural flexibility: This refers to the way overlapping participation links discussions across the network. We used social network analysis (SNA) to map connections, visualizing conversations as nodes (sized by number of posts and colored by recurrence). Iterative filtering retained first links with ≥3 and then ≥5 shared teachers, revealing the network's core, highlighting which conversations were central to the community's structure.

Results

Temporal flexibility: Teacher activity was not confined to institutional schedules. Nearly 71% of posts occurred during afternoon, evening, and night hours, demonstrating consistent engagement during personal time (Table 1). Active teachers contributed across all parts of the day, with only the late-night hours (after 21:59 until 6:59) showing lower activity, reflecting a natural boundary aligned with rest time. Some teachers used night hours, demonstrating how flexibility allowed engagement whenever individual schedules permitted.

Table 1

Distribution of WhatsApp Posts by Time of Day

Time of day	Posts (N)	Of total%	Percentage of teachers active in the community
School day (7:00-13:59)	682	29.6%	72.60%
Afternoon (14:00-17:59)	693	30%	68.49%
Evening (18:00-21:59)	825	35.8%	65.75%
Night (22:00-6:59)	107	4.6%	32.19%

N of posts = 2307; N of teachers =146

A little Over 40% (40.22%) of all posts were sent during school holidays and the summer break (Table 2). Post frequency during holidays (8.05 posts/day, average) was only slightly lower than on regular school days (8.67 posts/day, average), suggesting that the community remained active even when teachers were formally off duty. The summer break showed markedly higher activity (15.33 posts/day, average), highlighting the community's significance as a meaningful space for ongoing professional engagement beyond the school calendar.

Table 2

Distribution of Posts on Regular School Days and School Holidays.

Type of day	Posts (N)	Of total %	Days (N)
School day	1379	59.77%	159
Holidays	330	14.30%	41
Summer break	598	25.92%	39
Sum	2307	100%	238 (out of 365)

 © ISLS

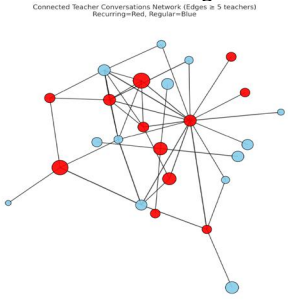
Sequential Flexibility Of the 349 conversations, 39 (11.2%) were recurring, comprising 585 posts (25.4% of the total). Although fewer in number, these conversations tended to be longer: on average, recurring conversations contained 15 posts compared to 5.6 posts in non-recurring conversations (Table 3). A Mann–Whitney *U* test confirmed that this difference was statistically significant, U (n_1 = 39, n_2 = 310) = 9544.50, z = 5.89, p < .001, r = .32. These findings indicate that recurring conversations accounted for a substantial share of interactions, reflecting teachers' sustained interest and motivation to extend professional dialogue. For example, a recurring discussion on reforms in humanities exams resurfaced three times, accumulating 40 posts, underscoring the role of recurring conversations in sustaining teacher engagement across time.

Table 3
Distribution of Recurring and Non-Recurring WhatsApp Conversations

Conversation type	Conversation (N)	Of total%	Posts (N)	Of total posts	Avg. posts per conversation
Non recurring	310	88.8%	1722	74.6%	5.6
Recurring	39	11.2%	585	25.4%	15.0
Total	349	100%	2307	100%	-

Structural flexibility: The social network analysis reveals the structural importance of recurring conversations. Out of 349 conversations, only 25 remained in the filtered network (edges ≥5 shared teachers), connected by 42 links. This indicates that only a small subset of discussions attracted broad participation across multiple teachers. The recurring conversations (red nodes) dominate the network's core, showing both larger node sizes (more posts) and denser connections (more teachers). Non-recurring conversations (blue nodes) are smaller and peripheral. Although recurring conversations accounted for only a small fraction of the total discussions, their nodes displayed disproportionately high betweenness centrality, functioning as bridging hubs that connected otherwise separate clusters in the network. This positioning highlights their structural role in sustaining the community's cohesion. This pattern suggests that recurring conversations are critical for integrating the community, linking different groups of teachers, and creating a cohesive space for professional dialogue (Figure 1).

Figure 1
Conversation Network (Edges ≥ 5 Shared Teachers). Red = Recurring, Blue = Non-Recurring. Node Size = Number of Posts.



Discussion and conclusion

This study demonstrates that flexibility in this WhatsApp-based community is a multifaceted practice manifested through temporal, sequential, and structural dimensions. Teachers consistently engaged beyond institutional schedules and revisited topics in recurring conversations that formed the core structure of the community network. Unlike flexibility in formal online courses - typically a predefined feature, the flexibility observed in the WhatsApp community was socially constructed, emerging from teachers' ongoing choices to initiate, sustain, and renew professional dialogue during their free time. Rather than viewing flexibility solely as a logistical feature, our findings highlight its enactment as a pedagogical practice, resonating with theoretical perspectives that emphasize its learning value. This study is an initial stage of a broader project. Subsequent analysis will expand on these findings, extend the categorization of message functions and content, and examine the relationship between voluntariness and flexibility. Taken together, this study highlights the socially constructed nature of flexibility and its role in supporting professional growth in digital environments.

 © ISLS

References

Barber, M., & Mourshed, M. (2007). *How the world's best-performing school systems come out on top*. McKinsey & Company.

Bragg, L. A., Walsh, C., & Heyeres, M. (2021). Successful design and delivery of online professional development for teachers: A systematic review of the literature. *Computers & Education*, 166, 104158. https://doi.org/10.1016/j.compedu.2021.104158

Bolam, R., McMahon, A., Stoll, L., Thomas, S., Wallace, M., Greenwood, A., ... & Smith, M. (2005). *Creating and sustaining effective professional learning communities* (Vol. 637). Research report. Universities of

Bristol, Bath, and London, Institute of Education. https://dera.ioe.ac.uk/id/eprint/5622/1/RR637.pdf

Collis, B., & Moonen, J. (2012). *Flexible learning in a digital world: Experiences and expectations*. Routledge.https://doi.org/10.4324/9780203046098

Fructuoso, I. N., Albó, L., & Beardsley, M. (2022). University students' preference for flexible teaching models that foster constructivist learning practices. *Australasian Journal of Educational Technology*, *38*(4), 22-39. https://orcid.org/0000-0002-6902-5818

Hymes, D. (1964). Introduction: toward ethnographies of communication 1. *American anthropologist*, 66(6_PART2), 1-34. https://doi.org/10.1525/aa.1964.66.suppl_3.02a00010

Huhang, L., Liang, M., Xiong, Y., Wu, X., and Lim C.P (2024) A systematic review of technology-enabled teacher professional development during COVID -19 pandemic. *Computers and education 223*.

Barber, M., & Mourshed, M. (2007). How the world's best performing systems come out on top.

Lantz-Andersson, A., Lundin, M., & Selwyn, N. (2018). Twenty years of online teacher communities: A systematic review of formally-organized and informally-developed professional learning groups. *Teaching and teacher education*, 75, 302-315. https://doi.org/10.1016/j.tate.2018.07.008

Lave, J., & Wenger, E. (1991). Situated learning: *Legitimate peripheral participation*. Cambridge University Press.

Lay, C. D., Allman, B., Cutri, R. M., & Kimmons, R.(2020). Examining a decade of research in online teacher professional development. In *Frontiers in Education* (Vol. 5, p. 573129). Frontiers Media SA. https://doi.org/10.3389/feduc.2020.573129

Lefstein, A., Snell, J., & Israeli, M. (2015). From moves to sequences: Expanding the unit of analysis in the study of classroom discourse. *British Educational Research Journal*, *41*(5), 866-885. https://doi.org/10.1002/berj.3164

Li, K. C., & Wong, B. Y. Y. (2018). Revisiting the definitions and implementation of flexible learning. *Innovations in open and flexible education*, 3-13.

Hadar, L. L., & Brody, D. L. (2017). Professional learning and development of teacher educators. *Handbook of Research on Teacher Education*, 1049-1064.

Loon, M. (2022). Flexible learning: a literature review 2016–2021. *Advance HE*.

Macià, M., & García, I. (2016). Informal online communities and networks as a source of teacher professional development: A review. *Teaching and teacher education*, *55*, 291-307.

Mourshed, M., Chijioke, C., & Barber, M. (2011). How the worlds most improved school systems keep getting better. *Voprosy obrazovaniya/educational studies Moscow*, (1), 7-25.

Scherz, Z., Eylon, B. S., & Yarden, A. (2021). Professional Learning Communities for Science and Technology Teachers: Models and Modeling. In *Long-term Research and Development in Science Education*, (pp. 290-305). Brill. https://doi.org/10.1163/9789004503625_013

Shulman, L. S. (2005). Signature pedagogies in the professions. *Daedalus*, *134*(3), 52-59.

Soffer, T., Kahan, T., & Nachmias, R. (2019). Patterns of students' utilization of flexibility in online academic courses and their relation to course achievement. *International Review of Research in Open and Distributed Learning*, 20(3). https://doi.org/10.19173/irrodl.v20i4.3949

Ulanday, M. L., Centeno, Z. J., Bayla, M. C., & Callanta, J. (2021). Flexible learning adaptabilities in the new normal: E-learning resources, digital meeting platforms, online learning systems and learning engagement. *Asian Journal of Distance Education*, *16*(2). https://orcid.org/0000-0002-4450-744X

Vangrieken, K., Meredith, C., Packer, T., & Kyndt, E. (2017). Teacher communities as a context for professional development: A systematic review. *Teaching and teacher education*, 61, 47-59. https://doi.org/10.1016/j.tate.2016.10.001

Vescio, V., Ross, D., & Adams, A. (2008). A review of research on the impact of professional learning communities on teaching practice and student learning. *Teaching and teacher education*, 24(1), 80-91.

Veletsianos, G., & Houlden, S. (2019). An analysis of flexible learning and flexibility over the last 40 years of Distance Education. Distance Education, 40(4), 454-468. https://doi.org/10.1080/01587919.2019.1681893

International Society of the Learning Sciences

Teachers' Evolving Sentiments and Competencies about Generative AI in Social Media Discourse: A Case Study on X

Karina Vázquez, Davinia Hernández-Leo, Shatha N. Alkhasawneh
karinaabigail.vazquez01@estudiant.upf.edu, davinia.hernandez-leo@upf.edu, shatha.alkhasawneh@upf.edu
Universitat Pompeu Fabra, Barcelona, Spain

Abstract: This study examines teachers' evolving sentiments and competencies regarding Generative Artificial Intelligence (GenAI) by analyzing teacher-authored posts on X (formerly Twitter) in 2023 and 2025. Combining qualitative content analysis with automated and manual sentiment analysis, the research is guided by the *UNESCO AI Competency Framework for Teachers (AI CFT)* and *Zhai's teacher agency framework*. The study contributes to the CSCL discourse by framing X as a space for informal, collaborative learning. Within this environment, educators share experiential insights and articulate evolving visions of GenAI as a collaborator in their pedagogical practice and professional development. Findings indicate a gradual shift toward more positive and practice-oriented engagement, though concerns about agency, accountability, and professional responsibility remain. A significant link between sentiment and competency suggests that emotional stance shapes readiness to adopt GenAI. The study informs professional development strategies aligned with *AI CFT curricular goals*, promoting ethical and human-centered GenAI integration in education.

Introduction and background

Generative Artificial Intelligence (GenAI) refers to deep learning models that generate original content based on patterns learned from large datasets (Sengar et al., 2025). Since the release of ChatGPT in 2022, its use in education has sparked both enthusiasm and concern. Teachers highlight its potential for personalized learning and efficiency, while raising issues related to critical thinking, problem-solving, and academic integrity (Ogurlu & Mossholder, 2023; Söderström et al., 2024). These developments are also reshaping teachers' roles and increasing demand for professional development, though how teachers' perspectives and human–AI collaboration practices have evolved remains unclear (Alkhasawneh et al., 2025; Cress & Kimmerle, 2023).

In this context, as during the COVID-19 pandemic (Beardsley et al., 2021), some teachers have turned to social media as CSCL spaces for informal, collaborative learning about GenAI. While survey studies show mixed attitudes toward GenAI, they are often limited in scope (DeVito et al., 2025). Platforms such as X (formerly Twitter) provide a large-scale and relatively unfiltered view of how teachers discuss and negotiate these technologies (Bruguera et al., 2019), building on its established role in peer learning and resource sharing (Orlanda-Ventayen & Ventayen, 2017; van Bommel et al., 2020). Recent studies indicate that teachers express both optimism and concern about ChatGPT, particularly regarding creativity, efficiency, plagiarism, and institutional pressures, with discourse becoming more nuanced over time (DeVito et al., 2025; Mamo et al., 2024; Naing & Udomwong, 2024).

This study analyzes teachers' evolving discourse on X in 2023 and 2025, combining sentiment analysis with two complementary theoretical perspectives*: the UNESCO AI Competency Framework for Teachers* (*AI CFT*; UNESCO, 2024) and *Zhai's (2024) teacher agency framework*. The *AI CFT* defines three progression levels—*Acquire* (awareness), *Deepen* (application and reflection), and *Create* (innovation and leadership)—operationalized through structured *learning objectives (LOs)* that capture teachers' development in knowledge, skills, and values. Although the framework includes five dimensions, this study focuses on the *Human-Centered Mindset*, as its *LOs* most directly reflect teachers' beliefs, attitudes, and values toward GenAI as expressed in discourse. *Zhai's framework* conceptualizes teacher agency through four roles—*Observer, Adopter, Collaborator*, and *Innovator*—highlighting evolving engagement and pedagogical integration while emphasizing that agency is dynamic and shaped by contextual factors.

Together, these frameworks provide a multidimensional lens to examine how teachers' sentiments, competencies, and roles evolve. From a CSCL perspective, GenAI is not only a tool but also a potential collaborator that reshapes agency and expertise in hybrid learning environments (Hernández-Leo et al., 2025). Accordingly, this study addresses the following research questions: (1) How have teachers' sentiments toward GenAI shifted in 2023 and 2025, as expressed in their discourse on X (formerly Twitter)?; (2) How do teachers' competencies, particularly within the *Human-Centered Mindset* dimension of the *AI CFT*, emerge and evolve in their social media discourse?; (3) How do these evolving sentiments and competencies reflect teachers' agency as conceptualized in *Zhai's (2024) framework* (*Observer, Adopter, Collaborator, and Innovator*)?; and (4) What

 © ISLS

professional development orientations emerge from *aligning AI CFT learning objectives* with *Zhai's* teacher roles?

Methodology

Data collection and preprocessing

This study draws on two datasets of social media posts retrieved from the Kaggle platform (Quaranta et al., 2021). To capture educators' discourse on GenAI, posts were filtered to include: (a) English-language content; (b) educator accounts identified through biography keywords (e.g., teacher, educator, professor); (c) education-related terms (e.g., student, class, school); and (d) references to ChatGPT as a proxy for GenAI, reflecting its prominence as the most widely recognized tool in public discourse. While this anchors the dataset in widely recognized discourse, it also limits the analysis to ChatGPT-related discussions. After filtering and deduplication, the dataset comprised 130 posts from 100 teachers in 2023 and 261 posts from 219 teachers in 2025, totaling 20,979 associated comments.

Teacher comments were coded using the *Human-Centered Mindset* dimension of the *UNESCO AI Competency Framework for Teachers* (*AI CFT*; UNESCO, 2024). A coding scheme based on its *learning objectives (LOs)* across the three progression levels—*Acquire*, *Deepen*, and *Create*—was applied. Each comment was independently coded by two trained coders in *ATLAS.ti*, a qualitative data analysis software (Hwang, 2007), using the framework's hierarchical notation (LO.X.Y.Z), where X denotes the *dimension*, Y the *progression level*, and Z the specific *learning objective*. Inter-coder reliability was ensured through independent coding followed by reconciliation discussions until consensus was reached.

To capture the affective dimension of teachers' engagement with GenAI, a triangulated sentiment analysis approach was employed (B. Liu, 2022), combining *VADER (Valence Aware Dictionary and sEntiment Reasoner)*, a rule-based model optimized for analyzing the informal and concise language typical of social media (Hutto, 2025; C. Hutto & Gilbert, 2014), and *RoBERTa (Robustly Optimized BERT Pretraining Approach)*, a transformer-based deep learning model capable of capturing more nuanced semantic and contextual patterns in text (Y. Liu et al., 2019), alongside manual classification. All comments were categorized as *positive*, *neutral*, or *negative*. Manual coding was prioritized in cases of disagreement between automated methods to capture contextual nuances such as sarcasm and irony.

Teachers' sentiments were analyzed alongside their AI-related competencies by linking each comment to relevant *UNESCO AI CFT learning objectives* (e.g., LO.1.1.1) and labeling it as *positive*, *neutral*, or *negative*. The datasets were merged in Python to produce a contingency table showing the distribution of *learning objectives* across sentiment categories. A Pearson's chi-square test of independence ($\alpha = 0.05$), implemented using the `scipy.stats.chi2_contingency` function in Python, was conducted to assess whether sentiment and competency levels were significantly associated with one another.

To examine teacher agency, *Zhai's (2024) framework* (*Observer*, *Adopter*, *Collaborator*, and *Innovator*) was applied to a filtered subset of 2025 comments previously coded at *AI CFT Level 2 (Deepen)*. From this subset, 64 comments containing explicit references to teachers' practices, attitudes, or use of GenAI were purposively selected and manually classified according to *Zhai's* roles. This process linked competency development with role progression and enabled a more nuanced understanding of teachers' engagement. By integrating the *AI CFT*'s competency structure (*Acquire*, *Deepen*, *Create*) with *Zhai's* socio-psychological perspective, the analysis captures both competency progression and evolving teacher roles in practice.

Finally, comments were coded according to *AI CFT learning objectives*, which were then linked to their corresponding *curricular goals* and aligned with *Zhai's roles*, forming a *learning objective–curricular goal–role* framework. This alignment captures competency progression alongside role-specific support needs and informs training implications. Coding was conducted independently by two researchers and subsequently reconciled through discussion to ensure reliability.

Results

This section reports (a) the distribution of teacher comments across *AI CFT Human-Centered Mindset* levels (*Acquire, Deepen, Create*) in 2023 and 2025; (b) year-to-year shifts in sentiment; (c) the association between competencies and sentiment; (d) the alignment of *AI CFT* competencies with *Zhai's role framework* using the 2025 *Deepen-level* subset; and (e) training implications derived from integrating *learning objectives, curricular goals*, and *roles*. Overall, the findings indicate a shift from initial awareness toward practice-oriented engagement, alongside persistent concerns about agency, accountability, and ethics.

 © ISLS

AI CFT coding across years

At the first level of the *AI CFT* framework (*Acquire*, LO.1.1), certain codes maintained consistent relevance across years (see Table 1). Comments reflecting on the benefits and limitations of GenAI persisted, indicating sustained foundational awareness among teachers. References to limitations increased slightly (9.2% to 10%), suggesting a shift toward more critical evaluation of tool functionalities. Meanwhile, references to human agency and the notion that AI is human-led declined markedly (10% to 3%), and discussions of human rights implications nearly disappeared. This pattern indicates a shift at this introductory level from foundational ethical awareness toward more practical, tool-oriented considerations.

Corresponding to the second level of the *AI CFT* framework (*Deepen*, LO.1.2), the data indicate a gradual pedagogical shift from apprehension toward emerging GenAI integration. While mentions of AI-enhanced methodology design (LO.1.2.4*)* increased modestly (1.5% in 2023 to 4.2% in 2025), concerns about the encroachment of AI on student cognition and teacher autonomy (LO.1.2.3) declined notably, suggesting growing professional confidence in human-centered AI use. However, engagement with institutional and policy-level dimensions remains limited. Although references to human supervision (LO.1.2.1) appear, broader considerations of regulatory frameworks and teacher accountability are largely absent, indicating a persistent gap in governance-related discourse.

At the third level of the *AI CFT* framework *(Create*, LO.1.3*)*, the data reveal a shift from general reflection toward more analytical engagement. While 50% of teachers' comments focused on reflecting on GenAI's educational implications in 2023, this proportion declined to 27.2% in 2025, as attention to critically evaluating those implications increased to 14.9%. Despite a slight rise in attention to mitigating educational risks, broader dimensions—such as GenAI's role in the workplace and its impact on citizenship development—remain limited. This pattern suggests that while teachers are moving toward higher-order engagement, important gaps persist in addressing social responsibility and policy-related aspects of GenAI adoption.

Table 1

Distribution of AI CFT Learning Objectives by Level in Teacher Comments (2023 and 2025).

AI CFT Level	Subcode	Learning Objective	2023	2025
Acquire (LO.1.1)	LO.1.1.1	Reflects on the benefits of AI tools	33.8%	32.6%
	LO.1.1.1	Reflects on the risks of AI tools	22.3%	6.1%
	LO.1.1.1	Reflects on the limitations of AI tools	9.2%	10.0%
	LO.1.1.2	Demonstrates awareness that AI is human-led	10.0%	3.0%
	LO.1.1.4	Understand the need to protect human agency	2.3%	1.1%
Deepen (LO.1.2)	LO.1.2.1	Understand the need for have human supervision in AI decisions	7%	3.8%
	LO.1.2.3	Predict the risk of AI usurping students' thinking process	12.3%	3.0%
	LO.1.2.3	Defends teachers' accountability	8.5%	1.9%
	LO.1.2.3	Demonstrate resistance without understanding	2.3%	6.5%
	LO.1.2.4	Designing methodologies that provide human interaction	1.5%	4.2%
Create (LO.1.3)	LO.1.3.1	Reflect on the implications of AI in education	50%	27.2%
	LO.1.3.1	Evaluate the implication of AI in education	4.6%	14.9%
	LO.1.3.1	Reflect on the implications of AI at work	6.0%	3.4%
	LO.1.3.2	Contribute to mitigating educational risks	7.7%	8.0%
	LO.1.3.3	Promote the development of citizenship qualities	0.8%	0.8%

Shifts in teachers' sentiment (2023 vs. 2025)

The sentiment analysis revealed notable shifts between 2023 and 2025. *Positive* sentiment rose from 36.15% to 40.23%, while *neutral* and *negative* attitudes declined to 40.61% and 19.16%, respectively. This trajectory suggests that increased familiarity with GenAI is gradually mitigating professional resistance. However, the persistence of *negative* sentiment in nearly one-fifth of the dataset indicates that substantial concerns remain, underscoring that while acceptance is growing, pedagogical and ethical apprehensions have not been fully resolved.

Sentiment analysis by AI CFT subcodes

The sentiment distribution across *AI CFT* subcodes reveals a clear association between pedagogical application and emotional stance. Subcodes focused on concrete practice, such as methodology design (LO.1.2.4) and risk mitigation (LO.1.3.2), were predominantly *positive*, suggesting increased teacher confidence in responsible integration. In contrast, LO.1.2.3 (encroachment and resistance) remained a primary source of *negative* sentiment, reflecting persistent concerns about the displacement of human functions. More abstract subcodes, such as AI authorship (LO.1.1.2) and general educational implications (LO.1.3.1), showed higher levels of *neutral* sentiment,

 © ISLS

indicating a more descriptive or cautious stance. Pearson's chi-square tests confirmed a significant association between sentiment and subcodes in both 2023 ($\chi^2 = 55.02$, $p < .001$) and 2025 ($\chi^2 = 98.74$, $p < .001$). This finding suggests that teacher attitudes vary by competency type, with stronger emotional responses linked to practical applications than to abstract considerations.

Integration of frameworks: Zhai's roles and AI CFT (Deepen Subset, 2025)

By mapping *Zhai's* teacher roles onto the *AI CFT Deepen* level (LO.1.2) using 2025 data, a pattern of progression from resistance to collaboration emerges. *Observers* are primarily linked to LO.1.2.3, reflecting limited trust and apprehension regarding AI encroachment. *Adopters* align with human supervision (LO.1.2.1) and tool accountability (LO.1.2.4), demonstrating a cautious, exploratory approach. In contrast, *Collaborators* are strongly associated with the design of AI-enhanced methodologies (LO.1.2.4), exhibiting high confidence and *positive* sentiment. The absence of *Innovators* suggests that transformative co-creation has not yet materialized. This integrated framework suggests that emotional responses influence role progression, highlighting the need for differentiated professional development strategies tailored to the specific curricular and attitudinal needs of each teacher profile.

Implications for teachers' training orientation

Aligning *AI CFT learning objectives*, *curricular goals*, and *Zhai's roles* (2025 *Deepen* level) reveals differentiated training needs. Comments linked to LO.1.2.3 were associated with *Observers*, who display caution and limited trust, indicating a need for foundational literacy, ethical framing, and peer exploration. LO.1.2.1 aligned with *Adopters*, who engage in cautious experimentation and benefit from practical use cases, mentoring, and supportive infrastructure. LO.1.2.4 was most often linked to *Collaborators*, who show readiness to integrate GenAI and benefit from co-design opportunities and dialogue on student agency. Overall, the findings highlight three key priorities for professional development: (a) the irreplaceable human role in education, (b) the importance of active supervision in GenAI adoption, and (c) the risks of over-reliance on GenAI tools without a deep understanding of their limitations. This alignment of *AI CFT curricular goals* with role-based development provides a structured, evidence-informed approach to designing differentiated training that is responsive to teachers' evolving needs.

Discussion and conclusion

This study examines the evolution of teacher discourse on X from 2023 to 2025, revealing a pedagogical shift from initial skepticism toward more practice-oriented engagement with GenAI (DeVito et al., 2025; Ogurlu & Mossholder, 2023). While year-to-year sentiment trends show modest growth in *positive* attitudes and a decline in *neutral* and *negative* stances, educators appear to be shifting from novelty toward more situated pedagogical uses (Mamo et al., 2024; Naing & Udomwong, 2024). Mapping these shifts to the *UNESCO AI CFT* and *Zhai's teacher agency framework* shows that while the *Acquire* level discussions remain stable, the *Deepen* and *Create* levels increasingly emphasize critical evaluation and risk-aware design. However, a persistent gap remains in policy and institutional accountability, reflecting prior findings that tool adoption often outpaces engagement with ethical frameworks (DeVito et al., 2025; UNESCO, 2024).

The integration of these frameworks clarifies how knowledge and attitudes evolve: *Observers* express resistance related to human replacement, *Adopters* engage in cautious experimentation focused on supervision, and *Collaborators* prioritize human-centered methodologies. The absence of *Innovators* suggests that transformative co-creation remains emergent and context-dependent (Hernández-Leo et al., 2025). Furthermore, the statistically significant association between sentiment and *AI CFT* subcodes indicates that *positive* affect aligns with confident, human-centered design, while *negative* sentiment reflects concerns about professional displacement.

These findings have important implications for teacher training. Following *UNESCO's curricular goals* for the *Human-Centered Mindset* dimension, professional development should: (a) reinforce the non-substitutable role of humans in educational processes, (b) model "human-in-the-loop" supervision, and (c) address the potential risks of relying heavily on GenAI tools without critically understanding their limitations (Ogurlu & Mossholder, 2023). Support should be role-sensitive: *Observers* require foundational literacy and ethical framing; *Adopters* benefit from practice-based mentoring and infrastructure; and *Collaborators* benefit from co-design opportunities and dialogue on student agency. This differentiated approach supports the balance between opportunities and risks while strengthening teacher agency (Hernández-Leo et al., 2025).

Limitations include the potential overrepresentation of vocal users on X, a narrow focus on ChatGPT, and the challenges of modeling sentiment in short texts. Future work should expand to other GenAI tools and more diverse educator communities—incorporating classroom artifacts and policy documents—while tracking longitudinal cohorts to identify sustained and responsible human–AI collaborations in education.

 © ISLS

References

Alkhasawneh, S. N., Lampropoulos, G., & Hernández-Leo, D. (2025). Generative AI in learning design: A systematic review. In International Conference on Artificial Intelligence in Education (pp. 17-30). Springer, Cham.

Beardsley, M., Albó, L., Aragón, P., & Hernández-Leo, D. (2021). Emergency education effects on teacher abilities and motivation to use digital technologies. *British Journal of Educational Technology*, 52(4), 1455-1477.

Bruguera, C., Guitert, M., & Romeu, T. (2019). Social media and professional development: A systematic review. *Research in Learning Technology*, *27*. https://doi.org/10.25304/rlt.v27.2286

Cress, U., & Kimmerle, J. (2023). Co-constructing knowledge with generative AI tools: Reflections from a CSCL perspective. *International Journal of Computer-Supported Collaborative Learning*, 18(4), 607-614.

DeVito, P., Vallala, A., Mcmahon, S., Hinda, Y., Thaw, B., Zhuang, H., & Kalva, H. (2025). *Unpacking Generative AI in Education: Computational Modeling of Teacher and Student Perspectives in Social Media Discourse* (No. arXiv:2506.16412). arXiv. https://doi.org/10.48550/arXiv.2506.16412

Hernández-Leo, D., Theophilou, E., Oshima, J., Minematsu, T., Matsueda, K., Naganuma, S., Cress, U., Chen, B., Zhu, X., del Castillo, F. D., Chen, W., Lyu, Q., Chen, X., Zheng, L., & Weinberger, A. (2025). *Collaborating with Generative AI for Learning?* In Proceedings of the 18th International Conference on Computer-Supported Collaborative Learning-CSCL 2025, pp. 525-533. International Society of the Learning Sciences. https://doi.org/10.22318/cscl2025.882858

Hutto, C., & Gilbert, E. (2014). VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *Proceedings of the International AAAI Conference on Web and Social Media*, *8*(1), 216–225. https://doi.org/10.1609/icwsm.v8i1.14550

Hutto, C. J. (2025). *Cjhutto/vaderSentiment* [Python]. https://github.com/cjhutto/vaderSentiment (Original work published 2014)

Hwang, S. (2007). Utilizing Qualitative Data Analysis Software. *Social Science Computer Review*. https://doi.org/10.1177/0894439307312485

Liu, B. (2022). *Sentiment Analysis and Opinion Mining*. Springer Nature.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). *RoBERTa: A Robustly Optimized BERT Pretraining Approach* (No. arXiv:1907.11692). arXiv. https://doi.org/10.48550/arXiv.1907.11692

Mamo, Y., Crompton, H., Burke, D., & Nickel, C. (2024). Higher Education Faculty Perceptions of ChatGPT and the Influencing Factors: A Sentiment Analysis of X. *TechTrends*, *68*(3), 520–534. https://doi.org/10.1007/s11528-024-00954-1

Naing, S. Z. S., & Udomwong, P. (2024). Public Opinions on ChatGPT: An Analysis of Reddit Discussions by Using Sentiment Analysis, Topic Modeling, and SWOT Analysis. *Data Intelligence*, *6*(2), 344–374. https://doi.org/10.1162/dint_a_00250

Ogurlu, U., & Mossholder, J. (2023). The Perception of ChatGPT among Educators: Preliminary Findings. *Research in Social Sciences and Technology*, *8*(4), 196–215. https://doi.org/10.46303/ressat.2023.39

Orlanda-Ventayen, C. C., & Ventayen, R. J. M. (2017). *Role of Social Media in Education: A Teachers' Perspective* (SSRN Scholarly Paper No. 3275172). Social Science Research Network. https://papers.ssrn.com/abstract=3275172

Quaranta, L., Calefato, F., & Lanubile, F. (2021). KGTorrent: A Dataset of Python Jupyter Notebooks from Kaggle. *2021 IEEE/ACM 18th International Conference on Mining Software Repositories (MSR)*, 550–554. https://doi.org/10.1109/MSR52588.2021.00072

Sengar, S. S., Hasan, A. B., Kumar, S., & Carroll, F. (2025). Generative artificial intelligence: A systematic review and applications. *Multimedia Tools and Applications*, *84*(21), 23661–23700. https://doi.org/10.1007/s11042-024-20016-1

Söderström, U., Hedström, E., Lambertsson, K., & Mejtoft, T. (2024). ChatGPT in education: Teachers' and Students' views. *Proceedings of the European Conference on Cognitive Ergonomics 2024*, 1–10. https://doi.org/10.1145/3673805.3673828

UNESCO. (2024). *AI competency framework for teachers.* United Nations Educational, Scientific and Cultural Organization. https://doi.org/10.54675/ZJTE2084

Van Bommel, J., Randahl, A.-C., Liljekvist, Y., & Ruthven, K. (2020). Tracing teachers' transformation of knowledge in social media. *Teaching and Teacher Education*, *87*, 102958. https://doi.org/10.1016/j.tate.2019.102958

Zhai, X. (2024). *Transforming Teachers' Roles and Agencies in the Era of Generative AI: Perceptions, Acceptance, Knowledge, and Practices* Journal of Science Education and Technology, 1-11.

Acknowledgements

This work has been supported by PID2023-146692OB-C33 and 2023-26 CEX2021-001195-M granted by MICIU/AEI/10.13039/501100011033 and KA220-HED-D8B72E6A. D. Hernández-Leo (Serra Húnter) also acknowledges the support of the ICREA Academia programme.

International Society of the Learning Sciences

Advancing Dialogic Leadership through AI-Enabled Training: Insights from a Design-Based Research Project

Xuanning Chen, Alliance Manchester Business School, University of Manchester, xuanning.chen@manchester.ac.uk
Imogen Casebourne, DEFI, Hughes Hall, University of Cambridge, ic407@hughes.cam.ac.uk
Rupert Wegerif, Faculty of Education, University of Cambridge, rw583@cam.ac.uk
Pedro Salinas, Yeltic, pedro@yeltic.com

Abstract: This paper reports insights from an ongoing design-based research project on developing an AI-enabled programme for dialogic leadership training. To ensure the effectiveness of the programme, one central challenge must be addressed: Can AI be designed to follow dialogic principles and provide trustworthy, supportive guidance that enables leaders to embed these principles into their practices? To address this, three design strategies were implemented: applying principles of dialogic intelligence and conversational heuristics to enhance AI's capacity for generative dialogue; integrating open-source datasets and psychological frameworks to improve AI's contextual awareness; and refining human-like interaction features to build AI's credibility. Guided by these strategies, an initial prototype for leadership in performance reviews was developed and tested with 20 participants, yielding positive results. Looking forward, the project will improve scenario realism, introduce scaffolding mechanisms and develop evaluation mechanisms for a better learning experience.

Introduction

Leadership plays a crucial role in shaping the value and culture of organisations (Ciulla, 2020). To be effective, leadership must be understood in relation to context and grounded in relational awareness (Cunliffe & Eriksen, 2011; Fletcher, 2012). Recent shifts in social contexts have created a pressing need for new forms of leadership. In particular, the contexts in which leadership operates have moved from tame with relatively simple and solvable problems, to wicked that characterised by heightened complexity, uncertainty, and full of problems with no definitive solutions (Bocala & Yurkofsky, 2025; Majee et al., 2025). Leaders now face the challenge of reconfiguring workplaces, retaining talent, and responding to rapidly changing market demands (Majee et al., 2025). Against this backdrop, there is a growing call for dialogic forms of leadership that can navigate complexity, build resilience, and capitalise on emerging opportunities (Ahmed et al., 2023; Kith, 2024).

The development of dialogic forms of leadership requires training programmes that are flexible, contextually grounded, and capable of embedding dialogic principles into the specific practical realities of leaders (Cunliffe & Eriksen, 2011; Kedian et al., 2016). Realising such training on a broad scale requires harnessing the potential of AI. Human trainers, mostly external to the organisation, can often face difficulties in designing training that is directly and consistently relevant to leaders' day-to-day practice (Kedian et al., 2016); also, hierarchical dynamics and authority concerns can discourage leaders from engaging openly with facilitators and training processes (Amestoy et al., 2014). AI, by contrast, is typically perceived as neutral and less judgmental (Borge et al., 2024), offering psychologically safe and adaptive environments that promote fuller participation. Besides, systems built on foundational models can interpret meaning in parallel, handle complexity, and adapt responsively to real-world scenarios (Alexander, 2024). Taken together, these capacities position AI not merely as a supplement but as a potentially significant enabler of leadership training.

However, delivering AI-based dialogic leadership training is not straightforward. A central hindrance is that AI often exhibits social sycophancy, accommodating users' preferences rather than constructively challenging them and supporting exploration of the unknown (Sharma et al., 2024). If unmitigated, this tendency risks reinforcing echo chambers and pushing leaders into monologic modes of leadership. To address this challenge, a design-based research project was undertaken by University of Cambridge and Yeltic, seeking to explore design strategies to enable AI to follow dialogic principles and support leaders to embed these principles into practice.

Design concepts and process

To ensure the design could be rapidly adapted as problems were encountered, the project followed agile principles, working in sprints to structure development (Banfield et al., 2015). In line with agile principles, prompt engineering was adopted as the central design method, as it enables rapid iteration, supports flexible adaptation to context, and allows continuous improvement through natural-language feedback (Bose et al., 2024; Kale et al., 2024). By leveraging the strengths of prompt engineering, academic and industry members of the team were able

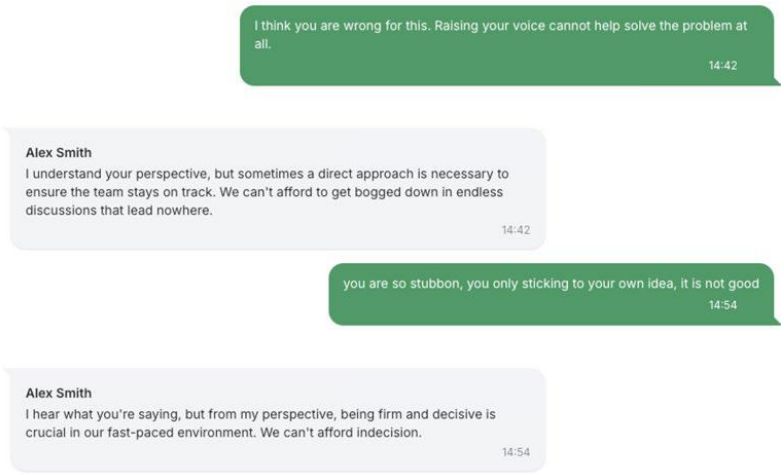
 © ISLS

to exchange ideas smoothly and respond effectively to emerging challenges, ensuring efficient collaboration and productive development processes that aligned with agile principles and harnessed the team's full potential.

Three groups of insights were adopted to guide the design of prompts intended to enable the AI to follow dialogic principles and provide trustworthy, supportive guidance for leaders in ways that are pedagogically effective and cost-efficient (i.e., minimising token usage):

(1) Central to the prompt design were dialogic intelligence principles (Wegerif, 2007; Wegerif, 2025) and heuristics for dealing with workplace conversations (Brown, 2013; Stone et al., 2023). These informed the development of prompts that enabled the AI to assume the role of a dialogic challenger, to mitigate its tendency toward social sycophancy and to nudge leaders to adopt a dialogic mode of leadership. After all, adults typically operate within established frames of reference and may regard their understanding as sufficient; only when confronted with dissonance or conflicts do they fully engage in the kind of learning that most effectively promotes growth (Mezirow, 1990). Figure 1 presents an example exchange involving an "unopen" AI agent under this prompt design.

Figure 1
Un "unopen" AI agent



(2) Datasets about workplace conversations and the discourse of individuals with different personality traits, systematically collected from open-source community platforms such as GitHub and Hugging Face. These datasets informed the development of prompts that assigned the AI contextually appropriate personas (e.g. distinct traits, conversation pattern, job seniority, and demographic backgrounds), to better mirroring the situations that leaders will typically encounter in their daily work.

(3) Insights from academic literature and professional training manuals, ensured that the prompts were pedagogically sound, theoretically grounded, and aligned with best practices in leadership development.

Guided by the design strategies outlined above, the first sprint of the AI-based platform was launched. Central to this sprint was a role-play session facilitated by an AI learning partner. The session focused on scenarios that frequently trouble leaders and demand sustained dialogic leadership, i.e. the performance reviews (Bradley & Campbell, 2016). Within this role-play, the AI partner adopted the role of a dialogic challenger, simulating behaviours commonly encountered in such contexts and encouraging leaders to practise dialogic engagement.

This first sprint was evaluated by the project team comprising education researchers, AI experts, business experts, and corporate managers, with the aim of identifying strengths and limitations in order to refine the version for prototype that could be tested with a more heterogeneous participant group.

The evaluation showed that the AI partner effectively addressed issues of social sycophancy. In addition, the training scenario and the AI partner's role were both judged highly relevant to managerial practice. Yet one central limitation was identified: the AI lacked human-likeness. This matters because perceived human-likeness shapes individuals' trust and engagement with AI (Ngo, 2025), which, in turn, can ultimately affect the effectiveness of training. To address this, strategies for enhancing human-likeness were introduced, including:

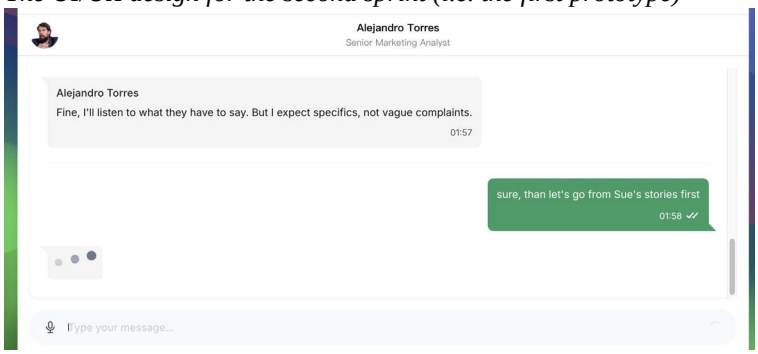
(1) The Big Five personality model and Plutchik's wheel of emotion (Mondal & Gokhale, 2020) were drawn on to enrich the prompts used to define the profile of the AI partner. The Big Five informed the design of new prompt blocks such as {behavioural_patterns}, which specified the AI partner's reaction modes and response tendencies linked to its assigned personality traits (e.g., agreeableness prompting

cooperative language, neuroticism prompting heightened sensitivity). Plutchik's wheel of emotion informed the creation of {emotional_states}, a prompt block assigning the AI partner dynamic, context-sensitive affective responses (e.g., anger, joy, anticipation) in alignment with its personality traits. Inclusion of these frameworks enabled the AI partner to be prompted to deliver emotionally attuned responses and present a more coherent personality.

(2) The refinement of UI/UX design (see Figure 2). More specifically, an avatar was added to the AI partner to strengthen social presence. These changes provided a more concrete and human-like impression, enhancing users' sense of interpersonal engagement. Also, the AI partner's response speed was adjusted, and behavioural indicators like "⋯" were incorporated to signal that the AI was "thinking." These modifications increased the perceived naturalness of the interaction. Finally, overall UI was redesigned to better mimic commonly used communication platforms to restore a real-life liked context.

Figure 2

The UI/UX design for the second sprint (i.e. the first prototype)



This refinement led to the second sprint of the training platform, which was again tested within the project team. The internal testing results indicated that the enhancements in human-likeness improved both authenticity and engagement. On this basis, the second sprint was integrated into the first prototype prepared for user testing.

User testing and findings

User testing of the initial prototype was conducted from 18th June to 25th July 2025. Twenty participants took part in the testing. Participants were selected to ensure a relatively balanced representation of professional status (i.e. 13 professionals with managerial experience and 7 students with background in either AI, dialogic education, or leadership) and gender (i.e. 12 females and 8 males) to support the validity of the findings.

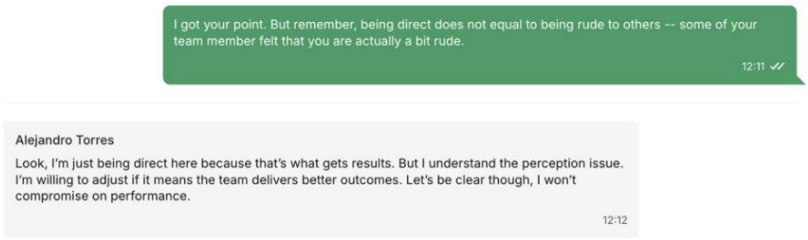
User testing took place in three stages. First, participants were introduced to the scenario, their assigned role, the role of the AI partner, and the goals and ground rules of the role-play. Next, participants interacted with the AI partner freely until the role-play goals were achieved. The average duration of interaction was around 20 minutes per session, with all histories recorded. Finally, participants reflected on their experiences and evaluated the design through a semi-structured interview and a follow-up survey, with focus on: (1) the AI partner's human-likeness and its ability to overcome social sycophancy; (2) the realism of the designed scenarios; (3) the applicability of leadership skills and dialogic thinking in practice; and (4) suggestions to refine the design.

The resulting data (i.e. interaction records and participant feedback) were analysed using thematic analysis supported by descriptive statistics. Analysis focused on three issues: (1) whether the AI partner behaved like a human actor capable of offering constructive challenge; (2) whether the designed scenarios reflected realities of workplace practice; and (3) whether the role-play supported development of dialogic leadership skills.

Findings indicate that all participants (n=20) agreed that the AI partner successfully emulated difficult human counterparts. The non-sycophantic behaviour of the AI partner is exemplified in Figure 3. Rather than offering encouragement or deferring to the participant's framing, the AI partner rejected the participants' response even when the participant had provided a reasonable justification, pressing participants to deepen their reasoning.

Figure 3

Participants-AI interaction history



These conversations, along with the scenarios in which they took place, were closely aligned with workplace realities for the professionals (n=13) and the internship experience of some students (n=3), as evidenced by participant feedback noting that:

> "This person is awful. I definitely see a lot of similarities with people in my workplace" (P9).
>
> "I would say this scenario can be frequently seen in my work" (P15).

Most participants (n=15) also confirmed that interaction with the AI partner supported the development of leadership skills and dialogic thinking. As P17 and P19, both as senior managers in their organisations, noted that the role-play helped them better practise leadership in realistic settings:

> "… as effective as a stress simulation with a human coach on soft skills in an MBA class" (P17).
>
> "This would have excellent training application for new HR professionals and new or existing line managers" (P19).

Also, interactions with AI partner provided a context for all participants (n=20) to explore how to apply core dialogic thinking in leadership settings. This was evident in their answers to the question, "*What have you learned from the role-play session?*" For this question, participants' answers consistently referred to core aspects of dialogic thinking like the following:

- Moving beyond fixed positions: "appreciating difference" (P1); "be patient and open" (P4).
- Deepening mutual engagement: "need to invite others to share ideas" (P6); "practising some strategies such as inviting others to think and talk, revoicing their words" (P7).
- Contingent reasoning: "a sentence can have different meaning that should figure out" (P12); "need to explain consequence and reason of one's actions" (P15).

Overall, findings from this initial user test indicate that the design strategies are feasible and appear to be pedagogically valuable, with the potential to establish a foundation for scalable AI-enabled dialogic leadership training.

Conclusion and next steps

This ongoing project represents an initial exploratory step toward understanding how AI might act as a partner for dialogic leadership development. Initial user testing confirmed the effectiveness of three design strategies: (1) structuring prompts using dialogic intelligence principles and conversational heuristics to reduce the social sycophancy of AI to guide reflective human-AI interaction; (2) leveraging open-source datasets and psychological frameworks to define the AI's personality and behavioral patterns to enhance the realism of human-AI interaction; and (3) integrating human-like UI/UX features such as conversational pacing and a human-like avatar to enhance naturalness, credibility, and sustained engagement.

The current design has several limitations. The most notable one is that the design was tested with only a single scenario and a relatively small group of participants. Given this, while retrieval-based grounding improves contextual accuracy, the system may still not fully capture the complexity of real-life leadership contexts, where nonverbal cues, relational history, and organizational stakes play a significant role. Also, the AI partner operated in a reactive way, responding to participants input rather than intervening proactively at critical moments.

As next steps and to address these limitations, the project team will broaden scenario coverage and conduct testing within organizations and engage larger, more diverse participant groups, aiming to acquire more user feedback, gain more insights about organizational context and simulate more realistic scenarios. The team will also introduce scaffold fading mechanisms, whereby the AI's challenger intensity is progressively reduced as learners demonstrate greater dialogic competence, supporting gradual role internalization rather than sustained reliance on AI provocation. Finally, a log-based assessment approach involving turn-level coding of dialogic moves will be developed to evaluate the depth and development of dialogic thinking over time, with a longer-term aim to explore how such training might contribute to preparing dialogically intelligent leaders who support sustainable and inclusive practices.

References

Ahmed, J. S., Masri, I. S., Minshar, M. M. A., & Abbas, A. F. (2023). The mediation role of psychological empowerment on the relationship between dialogical leadership and organizational brilliance. *Employee responsibilities and rights journal*, *35*(2), 227-246. DOI: 10.1016/j.edurev.2024.100618.

Alexander, K. (2024). The analysis of the efficiency of generative AI algorithms for creating a natural dialogue. *The American Journal of Interdisciplinary Innovations and Research*, *6*(06), 26-34. DOI: 10.37547/tajiir/Volume06Issue06-08 PAGE NO.: - 26-34.

Amestoy, S. C., Backes, V. M. S., Thofehrn, M. B., Martini, J. G., Meirelles, B. H. S., & Trindade, L. L. (2014). Dialogic leadership: strategies for application in the hospital environment. *Invest Educ Enferm*, 32(1), 119-127. DOI: 10.17533/udea.iee.v32n1a14.

Banfield, R., Lombardo, C. T., & Wax, T. (2015). *Design sprint: A practical guidebook for building great digital products*. O'Reilly Media.

Bocala, C., & Yurkofsky, M. (2025). Continuous improvement and the 'wicked'problems of racial inequity. *Journal of Educational Change*, *26*(2), 173-200. DOI: 10.1007/s10833-024-09520-1.

Borge, M., Smith, B. K., & Aldemir, T. (2024). Using generative ai as a simulation to support higher-order thinking. *International Journal of Computer-Supported Collaborative Learning*, *19*(4), 479-532. DOI: 10.1007/s11412-024-09437-0.

Bose, S., Tyagi, B., Sharma, P., & Sharma, A. (2024). Prompt Engineering: An efficient way for Content Creation using Prompt Chaining. In *2024 IEEE 4th International Conference on ICT in Business Industry & Government (ICTBIG)* (pp. 1-6). IEEE.

Bradley, G. L., & Campbell, A. C. (2016). Managing difficult workplace conversations: Goals, strategies, and outcomes. *International Journal of Business Communication*, *53*(4), 443-464. DOI: 10.1177/2329488414525468.

Brown, D. (2013). *Designing together: The collaboration and conflict management handbook for creative professionals.* Pearson Education.

Ciulla, J. B. (2020). *The search for ethics in leadership, business, and beyond*. Springer.

Cunliffe, A. L., & Eriksen, M. (2011). Relational leadership. *Human relations*, *64*(11), 1425-1449. DOI: 10.1177/0018726711418388.

Fletcher, J. K. (2012). The relational practice of leadership. In M. Uhl-Bien & S. M. Ospina (Eds.), *Advancing relational leadership research: A dialogue among perspectives.* (pp. 83-106). Information Age Publishing.

Kale, A., Yaduka, T., Shaikh, T., Kaur, G., Pinjarkar, L., Agrawal, P., ... & Khanuja, H. K. (2024, August). Unveiling the power of AI prompt engineering: A comprehensive exploration. In *2024 10th International Conference on Electrical Energy Systems (ICEES)* (pp. 1-9). IEEE.

Kedian, J., Giles, D., Morrison, M., & Fletcher, M. (2016). Leadership development as a dialogic process: the rationale and concept of an international leadership institute. *International Journal of Leadership in Education*, *19*(2), 182-202. DOI: 10.1080/13603124.2014.997800.

Kith, S. R. (2024). Dialogic Inclusive Leadership. *Organization Development Review*, *56*(4), 35-45.

Majee, W., David, I., Hultine Massengale, S., Wolff Rippel, C., James, D., Haq, T., & Russell, A. (2025). Creating spaces for dialogue and learning to build collective empowerment: A qualitative study on community leadership development in Midwest America communities. *Community Development*, *56*(3), 376-393. DOI: g/10.1080/15575330.2024.2411733.

Mezirow, J. (1990). A transformation theory of adult learning. In *31st Adult Education Research Annual Conference Proceedings* (vol. 31, pp. 141-146). ERIC.

Mondal, A., & Gokhale, S. S. (2020). Mining emotions on Plutchik's wheel. In *2020 seventh international conference on social networks analysis, management and security (snams)* (pp. 1-6). IEEE.

Ngo, V. (2025). Humanizing AI for trust: the critical role of social presence in adoption. *AI & SOCIETY*, 1-17. DOI: 10.1007/s00146-025-02506-4.

Sharma, N., Liao, Q. V., & Xiao, Z. (2024). Generative echo chamber? effect of llm-powered search systems on diverse information seeking. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (pp. 1-17). DOI: 10.1145/3613904.3642459

Stone, D., Patton, B., & Heen, S. (2023). *Difficult conversations: How to discuss what matters most*. Penguin.

Wegerif, R. (2007). *Dialogic education and technology: Expanding the space of learning*. Springer US.

Wegerif, R. (2024). Afterword: Dialogic space. *Theory into practice*, 63(2), 239-250. DOI: 10.1080/00405841.2024.2309840.

Wegerif, R. (2025, July 17). *Dialogic Intelligence*. https://substack.com/home/post/p-168581847.

International Society of the Learning Sciences

Cultivating Boundary-Crossing STEM Competencies: Building Community and Embracing Other Ways of Knowing in Global Informal Learning

Danielle P. Espino, Kristina M. Lux, Seung B. Lee, Eric R. Hamilton
danielle.espino@pepperdine.edu, kristina.lux@pepperdine.edu, seung.lee@pepperdine.edu, eric.hamilton@pepperdine.edu
Pepperdine University

Abstract: This paper examines how adolescents develop boundary-crossing STEM competencies through participation in an informal, globally networked virtual learning program. Using post-assessment interviews and recorded sessions of participating youth in Brazil, Kenya, Mexico, and the United States, we focus on two key competencies that emerged: Embracing Other Ways of Knowing and Building Community. These two competences reflect students' growing ability to engage across cultural and geographic differences and to establish inclusive, collaborative practices of learning and working together. The findings underscore the promise of boundary-crossing, project-based environments for developing STEM identities and fostering the human-centered skills necessary for globally connected STEM learning and future collaboration.

Introduction

In a constantly shifting technological and global landscape, the skills necessary for success will involve more than disciplinary knowledge. Students must develop interpersonal skills that help them communicate across cultures, work collaboratively, and navigate complexity. The idea of boundary-crossing competences is introduced by Walker and Nocon (2007) as the ability to function competently in multiple contexts, building off the idea of networked expertise brought forward by Moll et al (Moll, Tapia, & Whitmore, 1993). The need to support and foster such efforts become important in preparing youth for meaningful participation in future careers not only in STEM, but the broader global community.

Frameworks for education centered on 21st-century learning emphasize skills such as innovation, collaboration, and critical thinking (Starr et al., 2022; Salas Pico, 2013). While these skills are emphasized, many despite widespread endorsement of these skills, many youth learners—particularly those from marginalized communities—enter the postsecondary landscape with limited preparation of understanding or exposure from their educational experience (Allen et al., 2025). Supplemental learning experiences from informal and international programs may offer complementary or alternative pathways that more effectively foster such skills through culturally diverse, real-world opportunities often overlooked in formal settings (Jaiswal et al., 2024).

In STEM learning, another key component for adequate preparation is identity formation. Research studies have shown that youth who view themselves as capable and belonging in STEM are more likely to persist in these fields (Carlone & Johnson, 2007; Herrera & Kovats, 2022). In their paper, Carlone and Johnson (2007) identify three interrelated components of STEM identity that guided their initial analysis based on prior work: competence, or knowledge and understanding content; performance, or demonstrating one's scientific skills; and recognition, or being acknowledged as capable by self and others as a science person. These components are shaped in part by students' access to supportive social contexts and meaningful engagement.

Informal STEM learning environments have been found to foster settings that encourage identity development through hands-on learning, collaborative problem solving, and greater freedom to pursue student interests (Reid-Griffin et al., 2023; Nalipay et al., 2024). These kinds of learning environments reduce pressure for performance which allow learners to take intellectual risks, work with peers outside traditional classroom structures, and engage in culturally responsive ways. When programs also involve colleagues from other places, especially internationally, students have the opportunity to develop intercultural competence, language flexibility, and appreciation for diverse knowledge systems.

This paper builds on the idea of boundary-crossing competences to aspects of STEM-focused boundary-crossing competencies, informed from an informal, STEM project-based online learning environment for adolescent youth. Six competences emerged from the data, with two specific boundary-crossing competencies that are the focus of this paper: (1) Embracing Other Ways of Knowing, which describes students' openness to and appreciation of diverse epistemologies and local knowledge systems and (2) Building Community, which encompasses the development of social trust, shared purpose, and collaborative resilience among international

 © ISLS

peers. These competencies not only reflect what students learned, but how they approached collaboration, valued one another's contributions, and developed as STEM learners in relational and cross-cultural ways.

Context and methods

This study examines a project environment called the International Community for Collaborative Content Creation (IC4), a virtual STEM learning initiative that virtually connected adolescents from both high-income and low-income countries. Participants primarily came from Brazil, Kenya, Mexico, and the United States, with additional representation from countries including Cameroon, Finland, India, Iran, Namibia, Nigeria, Singapore, and Uzbekistan. Adolescent youth participated through afterschool programs or nonprofit organizations, meeting regularly in person locally within their site and joining global video conference sessions—called "global meet-ups"—to meet and collaborate with their international counterparts. These sessions were recorded, and students were interviewed for post-assessments at the close of the U.S. academic year.

Of the approximately 100 learners who participated in IC4 between 2020 and 2022, this study focused on data that involved six high school students selected for their consistent and active engagement throughout the entire period. Table 1 provides an overview of these students' demographics, reflecting students based in Brazil, Kenya, Mexico, and the United States. Our dataset includes nine recorded global meet-ups and a set of post-program interviews conducted in 2021. A team of four researchers transcribed and open coded the data, refining shared codes through an iterative process. These codes were then grouped into themes representing competencies fostered through the learning environment. While six competences were identified, two competencies were identified for the focus of this initial paper: Embracing Other Ways of Knowing and Building Community.

Table 1

Key Students Informing the Study (Pseudonyms Used for Student Names)

	Country	Gender	Initial Grade Level
Amani	Kenya	Male	9th
Benny	USA	Male	9th
Cecilia	Brazil	Female	8th
Dominic	Mexico	Male	8th
Liza	Brazil	Female	8th
Mary	USA	Female	10th

Additionally, epistemic network analysis (ENA), an approach of quantitative ethnography (QE), was utilized to develop statistical visualizations to support the descriptions of these competences (Shaffer, 2017). ENA is able to show the strength of connections between key constructs or codes in the data identified from the initial qualitative analysis (Siebert-Evenstone et al., 2017). In this paper, ENA was used to show how an individual's student discourse from global meet-ups reflected the development of the identified competencies over time. The strength of co-occurences between the constructs is reflected by the strength of the line, with thicker lines indicating a stronger relationship. Further description of ENA and QE can be found in the researchers' other work which utilizes this as the primary approach for the study (Espino et al., 2019; Espino et al., 2020; Espino et al., 2022). In this study, the ENA models for individual students included 12 relevant constructs from an iterated, grounded codebook identifying behavioral characteristics in the meet-up data (shared here with limited description due to space limitations, but planned for a larger work): (Cultural) Competence, (Project) Creation, Learner Centric (prior knowledge), (Content) Description, Curiosity, Project Input, Connecting (applying new knowledge), Openness (desire to be vulnerable, open to new experiences), Engagement, Resilience, Social Support, and Uncertainty.

Results and discussion

Embracing other ways of knowing

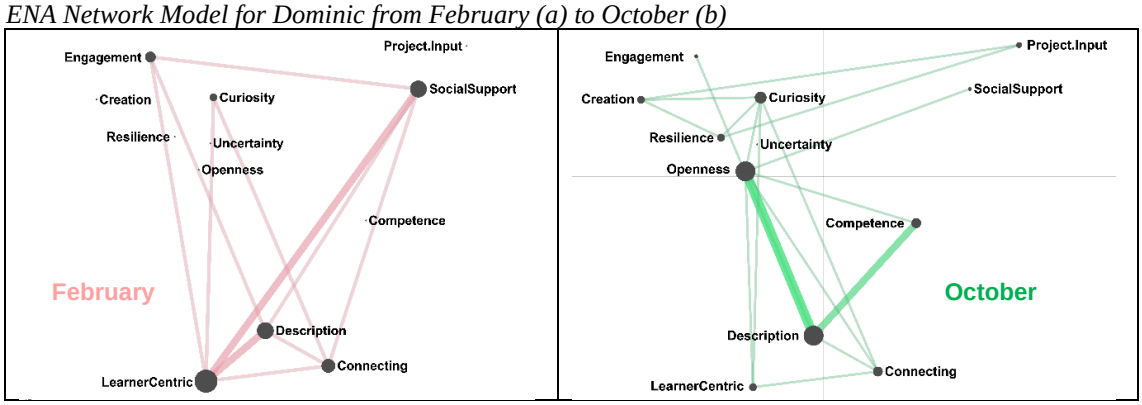
One of the most significant competencies identified was the ability to recognize and integrate diverse epistemologies. In global collaborations, students often encountered approaches to STEM that differed from their own. This kind of engagement parallels the concept of funds of knowledge (Moll et al., 2006), which asserts that culturally rooted knowledge systems hold value within learning environments. Similarly, Nasir and Hand (2008) highlight the importance of learners accessing and negotiating multiple knowledge traditions as part of identity development. In IC4, students did not simply tolerate difference—they came to appreciate it as central to learning.

Dominic, a student from Mexico, reflected on the mutual learning that occurred through collaboration. In a post-assessment interview, he emphasized the exchange of both cultural and academic knowledge: "We could help each other by using our own knowledge… I can share my culture with him." Dominic also noted how others'

expertise in mathematics helped expand his own understanding. His experience demonstrates how informal peer-to-peer learning across borders contributed to his developing STEM identity. This sentiment was also reflected in Dominic's discourse during global meet-ups over time, as seen in his ENA models in Figure 1 from February (left) to October (right). In February, there is not much acknowledgement of *Openness* or cultural *Competence*, but by October, Dominic is frequently relating content *Description* to both cultural *Competence* and a desire to be vulnerable to things he's not familiar with (*Openness*).

Figure 1

ENA Network Model for Dominic from February (a) to October (b)



(a) *(b)*

Amani, another student from Kenya, shared a story that illustrates both individual problem-solving effort and openness to support from an international collaborator. Initially working alone to resolve a coding issue, Amani described exhausting online resources before reaching out to a peer in the U.S., who offered direct support through a video call: "He told me it's as easy as saving... then he showed me how to use an autosave feature that saves a lot of time." This seemingly simple interaction highlights an important shift: from self-reliance to relational learning, rooted in trust and exchange across geographic and cultural boundaries. Amani also appreciated learning about global traditions during cultural presentations, recalling one in particular: "There was a student from Brazil with a presentation about some leaves used to make this beverage... I thought one day I should visit and perhaps have a taste." His curiosity to experience the beverage firsthand illustrates how STEM-related learning can intersect with cultural exploration, fostering both intellectual growth and global awareness.

These students exemplify how students moved beyond viewing STEM as a static knowledge skill set. Instead, they came to see knowledge as situated, culturally shaped, and worth engaging with on its own terms. The competence of Embracing Other Ways of Knowing thus reflects not just a collaborative practice, but a shift in how students see themselves and their place in STEM.

Building community

Students also developed competencies in creating inclusive, supportive, and collaborative learning communities. This competency involved more than simply joining a group; it required intentional actions—inviting participation, affirming contributions, and offering support. Students learned to take initiative in fostering psychological safety and belonging, even across language and time zone barriers.

Mary, a U.S. based-student reflecting further on her team's efforts, described how she and her partners found ways to connect despite language challenges. Mary and her team members had a willingness and openness to work through this together in order to reach a shared understanding of communication and effectively collaborate through their project. Through conversation, shared Slack messages, and even exchanging words in Portuguese, her team built trust and clarity. Mary describes this in detail during her interview:

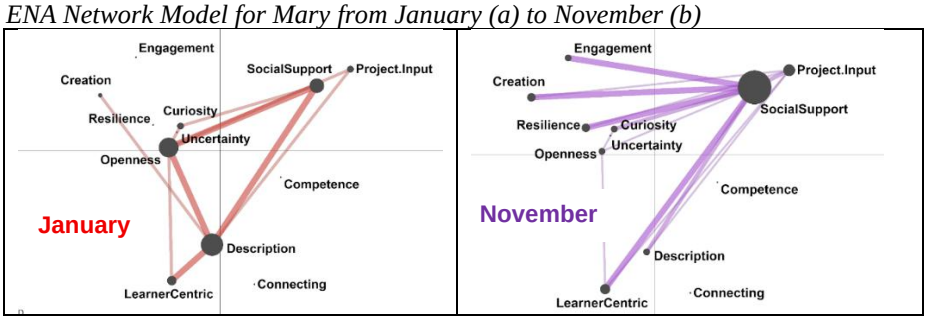
> "...all three of us [in the group] wanted just to get to know one another and actually talk with each other. A*lthough we may have a main language barrier but we found a way to talk…We, we talked about our ideas on the way we wanted to work on our projects, the sections that we each got, and the way that we want to do the work, and they were also nice in teaching a few words in Portuguese and go back and forth with them.* And then we also did text a lot through Slack, to just re-assure ourselves that we did understand what we were talking about and the way that we were going to do our project… with some of my peers."

This is also reflected in Mary's discourse during global meet-ups over time, as illustrated in her ENA models in Figure 2 from January to November. In January, Mary's discourse was primarily content-oriented

(*Description)*. Connections to *Social Support* were in support of content and supporting others with their lack of knowledge (*Uncertainty*). By November, Mary more prominently associates *Social Support* with *Engagement*, and *Resilience*, demonstrating how mutual encouragement and cross-cultural communication enabled her group to navigate language barriers, collaborate effectively, and sustain confidence throughout the project.

Figure 2

ENA Network Model for Mary from January (a) to November (b)



(a) *(b)*

Liza, a Brazilian student, spoke about how mutual openness and willingness to help were essential to completing group work. She appreciated how others consistently showed up, offered support, and accepted one another's contributions. Her post-assessment interview comments emphasized the importance of shared responsibility and care in making collaboration possible:

> *"Acho que sim acho que essa comparação com sendo aberto para pessoas que quiserem participar do projeto assim como a por todos e eu me ofereci para ajudar. Acho que isso ajuda muito na cooperação. A gente tem que estar aqui para que consigamos fazer um projeto juntos."* English Translation: *"I think so, I think this comparison with being open to people who want to participate in the project as well as that of everyone, and I offered to help. I think that helps a lot with cooperation. We have to be here for us to be able to do a project together and you can give a specific example of how you helped another club member or you were helped by another member of the club."*

In another individual reflection, Benny, a Mexican-American male learner from the United States, expressed how his consistent attendance at global meet-ups was not only for personal growth but also to support and encourage others, emphasizing a mutual commitment to the group. He discussed how the value of a non-hierarchical, welcoming space was important. Additionally, roles and credentials are secondary to shared curiosity and kindness:

> "If people, if new kids or teens or adults want to join… they don't have to be a very knowledgeable person. They don't have to be in the highest program in their school. They don't have to be the club president. They don't have to be the school president, don't have to be the principal to join because we will all learn from each other and only through kindness will we learn more about each other."

These accounts show that *Building Community* was not simply a product of the project's structure, but a competency developed through deliberate relational practices. Students learned to create environments where difference was welcomed, support was reciprocal, and everyone felt invited to contribute. In learning to navigate difference and foster belonging, students cultivated a critical competency for working across borders in STEM and beyond. Their experiences highlight that community building is both a social and intellectual endeavor—one that expands what it means to collaborate meaningfully in global contexts.

Conclusion

This study demonstrates how youth, when engaged in culturally diverse, informal STEM contexts, can develop boundary-crossing competencies vital for 21st-century collaboration. The competencies of *Embracing Other Ways of Knowing* and *Building Community* illuminate how adolescents build relationships, navigate complexity, and co-construct knowledge in ways that mirror the demands of global STEM careers. These findings point to the importance of educational designs that center human connection, relational learning, and cultural responsiveness. As educators and institutions seek to prepare learners for participation in international, interdisciplinary STEM fields, this study offers insight into the potential for young people to co-construct knowledge, bridge cultural boundaries, and reimagine collaboration as a deeply human endeavor.

References

Allen, T., Arriola, P. E., Breitenberger, C., Klyczek, K., Marrs, K. A., Matzner, S., Miller, K., Powell-Coffman, J., & Thatcher, M. (2025). Building Communities of Practice among Undergraduate STEM Departments to Foster Emergent Transformation: A Report on the Impact of Multiple-year Engagement within the PULSE Midwest and Great Plains Regional Network. *CBE—Life Sciences Education*, *24*(1), ar5.

Carlone, H. B., & Johnson, A. (2007). Understanding the science experiences of successful women of color: Science identity as an analytic lens. *Journal of Research in Science Teaching*, *44*(8), 1187–1218. https://doi.org/10.1002/tea.20237

Espino, D., Lee, S., Eagan, B., & Hamilton, E. (2019). An Initial Look at the Developing Culture of Online Global Meet-ups in Establishing a Collaborative, STEM Media-Making Community. In Lund, K., Niccolai, G. P., Lavoué, E., Hmelo-Silver, C., Gweon, G., & Baker, M. (Eds.), A Wide Lens: Combining Embodied, Enactive, Extended, and Embedded Learning in Collaborative Settings, 13th International Conference on Computer Supported Collaborative Learning (CSCL) 2019, Volume 2 (pp. 608-611). Lyon, France: International Society of the Learning Sciences.

Espino, D., Lee, S., Van Tress, L., Baker, T., & Hamilton, E. (2020). Analysis of U.S., Kenyan, and Finnish Discourse Patterns in a Cross-Cultural Digital Makerspace Learning Community Through the IBE-UNESCO Global Competences Framework. *Research in Social Sciences and Technology*, *5*(1), 86-100. https://doi.org/10.46303/ressat.05.01.5

Espino, D. P., Lee, S. B., Hokama, M., & Hamilton, E. R. (2022). Examining prompted discourse patterns in an informal, online, global collaborative learning environment. In Weinberger, A. Chen, W., Hernández-Leo, D., & Chen, B. (Eds.), Proceedings of the 15th International Conference on Computer-Supported Collaborative Learning - CSCL 2022 (pp. 179-186). International Society of the Learning Sciences.

Herrera, F., & Kovats Sánchez, G. (2022). Curando la comunidad [healing the community]: Community-centered STEM identity. *Journal of Hispanic Higher Education*, *21*(2), 135-150.

Jaiswal, A., Sapkota, M., Ramos, M. M., Starr, L., & Stahl, A. (2024). Nurturing Teamwork, Intercultural Competence, and Belonging in First-Year STEM Students. *International Journal for the Scholarship of Teaching and Learning*, *18*(2), 4.

Moll, L., Tapia, J., & Whitmore, K. (1993). Living knowledge: The social distribution of cultural resources for thinking. In G. Salomon (Ed.), Distributed cognitions: Psychological and educational considerations (pp. 1–46). Cambridge, MA: Cambridge University Press.

Moll, L., Amanti, C., Neff, D., & Gonzalez, N. (2006). Funds of knowledge for teaching: Using a qualitative approach to connect homes and classrooms. In *Funds of knowledge* (pp. 71-87). Routledge.

Nalipay, M. J., Huang, B., Jong, M. S. Y., Chai, C. S., & King, R. B. (2024). *Promoting STEM learning perseverance through recognizing communal goals: understanding the impact of empathy and citizenship. International Journal of STEM Education 11(17).*

Nasir, N. I. S., & Hand, V. (2008). From the court to the classroom: Opportunities for engagement, learning, and identity in basketball and classroom mathematics. *The Journal of the Learning Sciences*, *17*(2), 143-179.

Reid-Griffin, A., Croson, J., Fisher, S., & Lopez, N. (2023). Collaborative partnership: helping adolescents to develop a STEM mindset. *PDS Partners: Bridging Research to Practice*, *18*(1), 43-53.

Salas Pilco, S. Z. (2013). Evolution of the framework for 21st century competencies. *Knowledge Management & E-Learning: An International Journal*.

Shaffer, D. W. (2017). *Quantitative ethnography*. Madison, WI: Cathcart Press.

Siebert-Evenstone, A., Arastoopour Irgens, G., Collier, W., Swiecki, Z., Ruis, A. R., & Williamson Shaffer, D. (2017). In search of conversational grain size: Modelling semantic structure using moving stanza windows. *Journal of Learning Analytics*, 4(3), 123–139.

Starr, L., Yngve, K., & Jin, L. (2022). Intercultural competence outcomes of a STEM living–learning community. *International Journal of STEM Education*, *9*(1), 31.

Walker, D., & Nocon, H. (2007). Boundary-crossing competence: Theoretical considerations and educational design. *Mind, Culture, and Activity*, *14*(3), 178-195.

Acknowledgments

The authors acknowledge past funding support from the US National Science Foundation (Award # 2215613) for the work this paper reports. Views appearing in this paper do not reflect those of the funding agency.

International Society of the Learning Sciences

Age Differences in Cognitive Engagement and Their Association with Learning Performance in Collaborative Learning with Concept Maps: Experimental Investigation Using Epistemic Network Analysis

Shigen Shimojo, Yugo Hayashi
sshimojo@fc.ritsumei.ac.jp, yhayashi@fc.ritsumei.ac.jp
Ritsumeikan University, Osaka, Japan

Abstract: With the aging global population, lifelong learning is gaining attention. However, research on older adults' learning remains limited. This study reanalyzed collaborative learning with concept maps data from a previous study by applying the Interactive–Constructive–Active–Passive (ICAP)-based utterance analysis and network analysis to examine older adults' knowledge construction. By comparing them with young adults, we investigated the differences in learning processes and their effects on learning performance. Utterances were ICAP-classified and analyzed using epistemic network analysis. Young adults enhanced their learning by claiming their idea and reconstructing their understanding from others' knowledge (Constructive and Interactive). Conversely, older adults mainly engaged in recalling information (Active) and reflecting on a partner's concept map (Interactive). These findings clarify the mechanisms of older adults' learning from a theoretical computer-supported collaborative learning perspective and highlight the need for designs of support systems that compensate for Constructive and Active.

Introduction

As global aging accelerates, lifelong learning, particularly among older adults, is attracting increasing attention. According to a World Health Organization report (2025), the global population of older adults is progressively growing. In today's aging societies, learning is essential for extending working life and achieving social self-realization. Regarding older adults' cognitive characteristics, research has shown that fluid intelligence—knowledge-independent abilities such as critical thinking and processing speed—declines with age; conversely, crystallized intelligence, which includes knowledge and vocabulary, tends to increase and remains relatively stable even after peaking (Cattell, 1963). Against this backdrop, collaborative learning has the potential to serve as an effective strategy to promote learning without relying heavily on facilitators while helping compensate for cognitive decline.

In recent years, research on computer-supported collaborative learning (CSCL) has sought to identify collaborative learning processes that effectively promote learning and examine the types of scaffolding that foster deeper understanding among learners (Dillenbourg & Fischer, 2007). Group awareness enables learners to access and leverage others' knowledge as an external resource. Accordingly, it supports the development of mental models (internal representations of situations or concepts), construction of transactive memory (shared awareness of who holds what knowledge within a group), and formation of common ground (Clark, 1996). Therefore, to design CSCL environments that support older adults' learning, it is essential to capture the collaborative learning processes that emerge among them and clarify how learning occurs within those processes.

This study aimed to clarify how learning occurs in older adults during collaborative learning using concept maps. Few studies have examined the collaborative learning processes through which older adults' understanding deepens from the perspective of the constructive interaction theory. In our previous study (Shimojo & Hayashi, 2025) examined age-related differences between older and young adults in collaborative learning using concept maps. The results showed that young adults outperformed older adults in terms of learning performance. However, the previous analysis did not directly examine the relationship between detailed utterance-level processes and learning performance. Therefore, in the present study, we reanalyzed the same dataset (Shimojo & Hayashi, 2025) by integrating the Interactive–Constructive–Active–Passive (ICAP) framework (Chi & Wylie, 2014) with epistemic network analysis (ENA) to explicitly investigate how collaborative learning processes are associated with performance outcomes. Building on this approach, we combined fine-grained ICAP-based utterance analysis with network-based quantitative modeling to elucidate the knowledge construction processes of older adults during collaborative learning. The research questions were as follows:

RQ1: How different is the cognitive engagement of young and older adults during learning?

 © ISLS

RQ2: How does cognitive engagement affect learning performance?

Method

Participants and consent

A total of 52 individuals participated in the study: 26 university students majoring in psychology (10 males, 16 females; M = 19.77 years, SD = 1.90 years) and 26 older adults recruited from the Silver Human Resources Center (8 males, 18 females; M = 68.46 years, SD = 3.80 years). All older adults had completed university-level education, indicating broadly comparable formal educational backgrounds across groups. Although computer experience was not systematically measured, all participants demonstrated sufficient basic skills to operate the concept-mapping tool. The sample size was determined using G*Power, assuming a two-tailed t-test (d = 0.80, α = .05, power = .80), which indicated 26 participants per group. This study employed a one-factor (cohort: young vs. older adults) between-participants experimental design. The study received ethical approval from the Ethics Committee of Ritsumeikan University (Kinugasa-Hito-2022-22).

Experimental task

The objective of the learning task was to infer causal attributions and construct concept maps based on the attribution theory. Two types of experimental materials were used: a learning text and an episode. The learning task comprised an individual phase followed by a collaborative phase. Participants used a concept-mapping tool (Shimojo & Hayashi, 2020) that allowed simple creation and labeling of nodes and links; no advanced technical skills were required. All participants were novices in attribution theory and concept mapping. The learning text introduced the attribution theory, focusing on the causal explanations of success and failure. The episode material was a Japanese translation of the counseling session content reported by Weinberger et al. (2006), in which students expressed anxiety about the upcoming semester. During the 15-minute individual phase, participants read the text, inferred causes of the protagonist's anxiety, and independently constructed a concept map (e.g., anxiety–internal–effort). In the 20-minute collaborative phase, pairs explained their reasoning, referred to their individual maps (displayed on the right), and jointly created a new map (displayed on the left).

Experimental procedure

The learners first read a text on the attribution of success and failure. They then completed a pretest assessing their prior understanding, which asked, "Please describe what you know about the attribution of causes for success and failure." Following this, they engaged in the individual and collaborative phases. After completing the task, a posttest was administered to evaluate participants' understanding, which stated, "Based on the content discussed, describe what you know about the attribution of causes."

Measurement

In this study, two outcome measures were used: (1) the collaborative learning process, and (2) learning performance.

Collaborative learning process

Learners' utterances were coded using the ICAP framework into three categories: Active (repeating/summarizing text), Constructive (generating or elaborating ideas), and Interactive (building knowledge with a partner) based on Chi & Wylie (2014) and Wang et al. (2015). A primary coder scored all data, with 20% independently coded by a secondary coder (Krippendorff's α = .68). Although the reliability coefficient was slightly below the conventional threshold, all instances of disagreement were discussed between the coders until consensus was reached, and the coding scheme was refined where necessary. The agreed-upon codes were then reflected in the final dataset used for analysis. ENA was applied to visualize and quantify co-occurrence patterns among utterances. Connections were computed using a sliding-window of seven preceding utterances, and network structures were compared between young and older adults (GroupID). Principal components (MR1, SVD2) and connection strengths were calculated and statistically compared using the Benjamini–Hochberg FDR correction. All ENA analyses were conducted using the rENA package in R (Marquart et al., 2018).

Learning performance

Learners' understanding was assessed using pretest–posttest difference scores on an open-ended question about attribution theory. Responses were scored 0–5 based on the degree of causal attribution and explanation elaboration. A primary coder scored all responses, and 20% were independently coded by a secondary coder (Krippendorff's α = .91). Coders were blinded to conditions. Logistic regression was used to examine differences

in learning performance, with the difference score as the dependent variable and ENA dimensions and cohort as predictors.

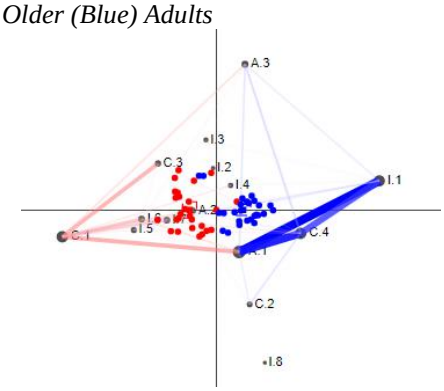
Findings

Differences in the collaborative learning process through the epistemic network analysis

This section addressed RQ1: How do cognitive engagement levels differ between young and older adults during learning? Figure 1 presents a difference graph comparing the scores of young (red) and older (blue) adults. Along the x-axis, a significant difference was observed between older (M = 0.30) and young (M = –0.28) adults, t (49.44) = –11.41, p < .001, Cohen's d = –2.95. By contrast, no significant difference was found along the y-axis between older (M = 0.00) and young (M = 0.00) adults, t (47.06) = 0.00, p = 1.00, Cohen's d = 0.00. The figure illustrates that older adults demonstrated strong connections between the Interactive and Active modes, whereas young adults demonstrated connections between the Interactive and Constructive modes. Additionally, we compared older and young adults using the Mann–Whitney U test to examine whether any two codes differed significantly in strength. In the comparisons, p-values were adjusted using the False Discovery Rate (FDR). We found that older adults exhibited significantly stronger connections for A-1–C-4 (U = 136.00, p = .006, r = .51), A-1–I-1 (U = 60.00, p < .001, r = .73), A-3–I-1 (U = 208.50, p = .017, r = .44), and C-4–I-1 (U = 90.00, p < .001, r = .68). By contrast, young adults showed stronger connections for A-1–I-6 (U = 479.00, p = .013, r = .47), C-1–C-3 (U = 499.00, p = .013, r = .47), and C-1–I-7 (U = 496.50, p = .014, r = .46). These findings indicate that young adults tend to engage more in Constructive and Interactive, while older adults emphasize Active and Interactive mode.

Figure 1

The Subtracted Network of Cognitive Engagement Between Young (Red) and Older (Blue) Adults



Effect of cognitive engagement on learning performance

This section addressed RQ2: How does cognitive engagement affect learning performance? We conducted a logistic regression analysis with improvement in learning performance as the dependent variable and the principal components, MR1 and SVD2, with age group (older vs. young adults), as the independent variables. The results indicated that MR1 and membership in the younger group were significantly associated with improved learning performance (Table 1). Therefore, higher MR1 scores and younger age increased the likelihood of performance gains. They suggested that the utterances contributing to improved learning performance were characterized by externalizing one's own opinions (C-1), confirming others' views, and reconsidering one's own thoughts, rather than relying on combinations of I-1 with questioning or recalling utterances.

Table 1

Logistic Regression Predicting Learning Performance Improvement

Variable	*B*	*SE*	*OR*
MR1	5.56*	2.46	260.04
SVD2	0.28	1.24	1.32
Group (Young)	4.92**	1.73	137.41

Note. N = 52. $-2log\lambda$ = 13.72, p = .003, AIC = 62.55. Group (Young) represents the younger group (a dummy variable, with the older group as the reference category). OR = odds ratio ($Exp(B)$). *p < .05, **p < .01.

 © ISLS

Discussion

We comprehensively examined the collaborative learning processes of older and young adults when using concept maps and investigated whether these processes influence learning performance. Specifically, we addressed two research questions:

RQ1: How different is the cognitive engagement of young and older adults during learning?

RQ2: How does cognitive engagement affect learning performance?

Differences in collaborative learning processes between young and older adults

In this section, we discussed how cognitive engagement differs between young and older adults during learning (RQ1). Previous studies on older adults' collaborative learning have demonstrated that cognitive abilities, such as memory, executive function, and processing speed, naturally decline with age (Salthouse et al., 2003). Moreover, the literature on aging has often treated memory as an independent and solitary phenomenon that has been studied individually (Mahr & Gsibra, 2018). By contrast, we examined how cognitive engagement manifests in older adults in the context of collaborative learning. Wolfe et al. (2023) emphasized that the outcomes of collaborative learning depend on the nature of the task. While several studies have focused on cooperative learning, where roles can be clearly divided, our study emphasized collaborative learning, where tasks are less easily partitioned and meaning must be co-constructed through interaction and dialogue.

These differences can be theoretically explained using the ICAP framework (Chi & Wylie, 2014). Young adults externalized their ideas (Constructive) and reexamined them through others' input (Interactive), reaching the interactive level, which can yield the most profound learning effects. Conversely, older adults experienced a more surface-level connection between the Active and Interactive levels, devoting their cognitive resources to recall and questioning rather than to co-constructing knowledge. This suggests that the socially shared regulation of learning (Hadwin et al., 2011) is more readily established among younger adults but poses greater challenges for older adults.

Influence of collaborative learning processes on learning performance

This section discussed how cognitive engagement affects learning performance (RQ2). Previous studies have shown that the outcomes of collaborative learning depend on the nature of the learning content (Wolfe et al., 2023). However, most studies have focused on cooperative learning, in which tasks can be divided and completed individually. From the perspective of the constructive interaction theory, few studies have examined the relationship between utterance processes and learning performance in collaborative learning using concept maps (Shimojo and Hayashi, 2025). Therefore, this study investigated them by combining ICAP-based utterance analysis with quantitative evaluation through network analysis (Shimojo and Hayashi, 2025).

This difference in how cognitive engagement patterns relate to learning performance can be interpreted from the perspective of transactive memory systems. Younger learners externalized their knowledge through C-1 and repeatedly engaged in mutual confirmation, thereby sharing "who knows what" and managing knowledge in a distributed manner across the group. This process appeared to strengthen collective learning outcomes. Furthermore, these findings support the importance of collaborative meaning making as emphasized in CSCL research (Roschelle & Teasley, 1995). Therefore, learning performance is promoted not by the mere recall of knowledge but by the dialogic processes in which externalized ideas are mutually confirmed, revised, and co-constructed.

Limitations and future work

The small number of ICAP utterances limits generalizability, and the sample may not represent typical young or older adults. The study design does not allow definitive causal claims about aging effects, and some large effect sizes and odds ratios may reflect model instability or overfitting due to the small sample size; thus, results should be interpreted cautiously. Although the tasks required no prior domain knowledge, it remains unclear whether they were truly age-neutral, as factors such as digital familiarity or processing speed may have differed across groups, limiting internal validity. In addition, although the learning outcome measure demonstrated high reliability ($\alpha = .91$), the mean scores were moderate (pretest: 2.25; posttest: 2.60), suggesting limited ceiling effects but also somewhat restricted score variance. This restriction may have constrained the ability to fully capture differences in learning gains.

Future studies should increase the sample size, extend task duration, and employ model-based approaches to examine underlying cognitive processes and supports for reducing age-related disadvantages. More precise analyses of how constructive engagement (e.g., C-1) can be facilitated and how cognitive engagement levels emerge from specific processes are also needed.

References

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology, 54*(1), 1–22. https://doi.org/10.1037/h0046743

Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist, 49*(4), 219–243. https://doi.org/10.1080/00461520.2014.965823

Clark, H. H. (1996). *Using language*. Cambridge University Press.

Crompton, C. J., Wolters, M. K., & MacPherson, S. E. (2022). Learning with friends and strangers: Partner familiarity does not improve collaborative learning performance in younger and older adults. *Memory, 30*(5), 636–649. https://doi.org/10.1080/09658211.2022.2041038

Dillenbourg, P., & Fischer, F. (2007). Computer-supported Nederland. collaborative learning: The basics. *Zeitschrift für Berufsund Wirtschaftspädagogik, 21*, 111–130.

Engelmann, T., & Hesse, F. W. (2010). How digital concept maps about the collaborators' knowledge influence computer-supported collaborative problem solving. *Computer-Supported Collaborative Learning, 5*(4), 299–319. https://doi.org/10.1007/s11412-010-9089-1

Hadwin, A. F., & Oshige, M. (2011). Self-regulation, coregulation, and socially shared regulation: Exploring perspectives of social in self-regulated learning theory. *Teachers College Record, 113*, 240-264. https://doi.org/10.1177/016146811111300204

Krippendorff, K. (2018). *Content analysis: An introduction to its methodology (4th ed.)*. Sage Publications.

Mahr, J. B., & Csibra, G. (2018). Why do we remember? The communicative function of episodic memory. *Behavioral and Brain Scien*ces, *41*, 1–93. https://doi.org/10.1017/S0140525X17000012

Marquart, C. L., Swiecki, Z., Collier, W., Eagan, B., Woodward, R., & Shaffer, D. W. (2018). rENA: Epistemic Network Analysis. R package version 0.1.5. Retrieved from https://cran.r-project.org/web/packages/rENA/index.html

Roschelle, J., & Teasley, S. D. (1995). The construction of shared knowledge in collaborative problem solving. In C. E. O'Malley (Ed.), *Computer supported collaborative learning* (pp. 69–97). Springer. https://doi.org/10.1007/978-3-642-85098-1_5

Ruis, A. R., Siebert-Evenstone, A. L., Pozen, R., Eagan, B., & Shaffer, D. W. (2019). Finding common ground: A method for measuring recent temporal context in analyses of complex, collaborative thinking. In K. Lund, G. Niccolai, E. Lavoué, C. Hmelo-Silver, G. Gwon, & M. Baker (Eds.), *A wide lens: Combining embodied, enactive, extended, and embedded learning in collaborative settings: 13th International Conference on Computer Supported Collaborative Learning (CSCL)* (pp. 136–143).

Salthouse, T. A., Atkinson, T. M., & Berish, D. E. (2003). Executive functioning as a potential mediator of age-related cognitive decline in normal adults. *Journal of Experimental Psychology General, 132* (4), 566–594. https://doi.org/10.1037/0096-3445.132.4.566

Shaffer, D. W., Hatfield, D., Svarovsky, G., Nash, P., Nulty, A., Bagley, E., Frank, K., Rupp, A., & Mislevy, R. (2009). Epistemic Network Analysis: A prototype for 21st century assessment of learning. *International Journal of Learning and Media, 1*(2), 33-53.

Shimojo, S., & Hayashi, Y. (2020). Prompting learner–learner collaborative learning for deeper interaction: Conversational analysis based on the ICAP framework. In *Proceedings of the 28th International Conference on Computers in Education (ICCE 2020)* (pp. 177–182).

Shimojo, S., & Hayashi, Y. (2025). Age-related differences in explanatory activities during collaborative learning with concept maps: Experimental investigation using epistemic network analysis. *International Journal of Computer-Supported Collaborative Learning, 20*, 491–518. https://doi.org/10.1007/s11412-025-09456-5

Wang, Y., Paquette, L., & Baker, R. S. (2015). A longitudinal study on learner career advancement in MOOCs. In C. Conati, N. Heffernan, A. Mitrovic, & M. F. Verdejo (Eds.), *Proceedings of the 8th International Conference on Educational Data Mining* (pp. 406–409).

Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & Education, 46*(1), 71–95. https://doi.org/10.1016/j.compedu.2005.04.003

Wolfe, K., Crompton, C. J., Hoffman, P., & MacPherson, S. E. (2023). Collaborative learning of new information in older age: A systematic review. *Royal Society Open Science, 10*(10), Article 211595. https://doi.org/10.1098/rsos.211595

World Health Organization (WHO). Dementia. Available online: https://www.who.int/news-room/fact-sheets/detail/dementia (accessed on 9 April 2025).

Acknowledgments

This study was supported by the JSPS KAKENHI (grant numbers JP23K28200 and 23H03510).

International Society of the Learning Sciences

From Idea Consumers to Knowledge Co-Creators: A 3-Year Design Study of Sustained Creative Work Through Cross-Community Partnership

Robert Huang, University of Toronto, rob.huang@utoronto.ca

Abstract: This three-year design-based research (DBR) study investigates how to shift students from being passive consumers of information to active co-creators of knowledge. Conducted in three iterative phases, the study progressed from a local Grade 2 classroom to a cross-national partnership between Grade 2 (Canada) and Grade 9 (Spain). Using a mixed-methods approach, including a novel AI-based analytical tool (StAS), results demonstrate a clear progression from students reproducing authoritative information to engaging in authentic knowledge synthesis and co-design. These findings provide a model for cultivating epistemic agency by progressively expanding the "public" for student ideas and utilizing multimodal tools for collaborative theory building.

Introduction and theoretical framework

At its core, Knowledge Building is defined by the production and continual improvement of ideas for the benefit of a community (Scardamalia & Bereiter, 2003). This process positions learners as epistemic agents: individuals who take responsibility for advancing a community's knowledge (Scardamalia, 2002). Yet a persistent challenge in education, and specifically within Computer-Supported Collaborative Learning (CSCL), is the gap between this vision and classroom reality. Students often remain in a "learning mode"—reproducing accepted facts from authoritative sources—rather than shifting to a "design mode," where they treat ideas as artifacts to be continually improved (Bereiter & Scardamalia, 2003). This distinction is critical. In "belief mode," typical of schooling, students treat ideas as personal possessions to be accepted or rejected based on authority. In contrast, "design mode" requires a shift in epistemic stance: ideas become public entities, distinct from the author, akin to a blueprint or a prototype. Just as engineers do not take personal offense when a prototype fails, students in design mode view the critique of an idea not as a personal attack, but as a necessary step in advancing community knowledge.

This paper documents a three-year design-based research (DBR) study that documents the iterative design of interventions aimed at fostering this capacity for sustained creative work. We argue that moving students from reproducing knowledge to co-creating it requires a carefully scaffolded approach that cultivates their epistemic agency. This is achieved by progressively expanding the "public" for their ideas, from the individual to the classroom and finally to a networked external community.

This scaffolded approach requires two key supports. First, an open and public community space, such as Knowledge Forum, facilitates discourse across geographical areas and allows for the emergence of shared goals (Gloor, 2006). Second, multimodality is essential for making ideas accessible and improvable. It allows creators to represent their thinking in generative and collaborative ways (Lewis, Pea, & Rosen, 2010), inviting others to build upon them in a manner that supports authentic learning and connection (Herrington & Kervin, 2007).

Methodology

Research approach and context

This study utilized a design-based research (DBR) approach (Brown, 1992) featuring iterative design interventions. The primary research site was a Grade 2 class at an independent school in Toronto, Canada. The curriculum focused on salmon life cycles, selected because it presented a complex ecological system requiring students to understand interdependence, human impact, and conservation efforts rather than memorize isolated facts. Study 3 expanded to include a second site with a Grade 9 class in Tarragona, Spain.

- Study 1 (Jan-Apr 2020): Baseline study establishing challenges in idea generation.
- Study 2 (Jan-May 2021): Implementation of a local, multimodal digital storytelling environment.
- Study 3 (Oct 2021-May 2022): Creation of a cross-national partnership (Grade 2 & Grade 9).

Each year involved a new group of approximately 22-23 Grade 2 students; Study 3 also included 60 Grade 9 students. The same Grade 2 teacher participated across all studies, providing crucial continuity.

 © ISLS

Data sources and analytical approach

A mixed-methods approach analyzed multiple data sources: (1) audio/video recordings of classroom activities, (2) student-generated notes within Knowledge Forum, and (3) student digital stories. Knowledge Forum is an electronic communal workspace designed to support Knowledge Building, allowing users to contribute notes, build on others' ideas via scaffolds (e.g., "I need to understand"), and organize inquiry through customizable views (Scardamalia, 2004).

Qualitative analysis utilized inductive coding (Charmaz, 2006). While conducted by a single coder, trustworthiness was established through prolonged engagement and peer debriefing to verify the alignment of codes with the theoretical framework.

AI-assisted content analysis (StAS)

To quantitatively track the shift from reproduction to creation, we utilized the "Similarity to Authoritative Sources" (StAS) tool (Khanlari, Huang & Scardamalia, 2024). This custom Python-based application uses a Sentence-BERT (SBERT) model to measure the semantic distance between student ideas and expert content. We established a consistent corpus of 10 expert-created videos on salmon life cycles. To validate the tool, we conducted empirical testing on the corpus itself. The average similarity between distinct concepts within the source material was 0.35, while direct paraphrases consistently scored above 0.70. Based on this, a similarity score of 0.5 was selected as a conservative threshold: notes scoring above 0.5 were classified as reproducing source material, while lower scores indicated the introduction of novel ideas or synthesis.

Findings

Study 1: Challenge of “design mode”

Study 1 revealed the difficulty of shifting students out of "learning mode." Students engaged in rich activities—raising salmon in a classroom hatchery, creating plasticine models, and hosting visiting experts. However, analysis of Knowledge Forum notes revealed that the digital environment was primarily used to record oft-repeated information from authoritative sources.

The StAS tool supported this observation. Many student notes had high similarity scores (average ~0.69). For example, notes such as "how do salmon eggs grow" (Score: 0.69) and "why do salmon have yolk sacs" (Score: 0.68) were near-direct paraphrases of the expert video transcripts. This highlighted the difficulty of getting students to practice self-regulatory processes, such as identifying "promising ideas" (Chen, Scardamalia, & Bereiter, 2015), iterating on plans based on new inputs, or identifying areas for further research.

Study 2: Multimodality and social connection

Study 2 introduced multimodal digital storytelling to address this. Students produced 18 digital stories, stitching together video and plasticine modeling to explain the salmon life cycle. While engagement increased, the artifacts often functioned as "end-of-term essays"—culminating products rather than improvable ideas. StAS scores remained high, indicating that while the medium had changed, the epistemic mode remained reproductive.

However, a crucial shift occurred in the social dynamics. Students began to exhibit characteristics important for sustained creative work: social-emotional connection (Immordino-Yang, 2016). Students posted congratulatory build-ons to convey feelings upon watching peers' videos:

> "Love your video... I didn't know that they eat that kind of food."
>
> "Wow that video taught me a lot Thanks."

This suggests that a Knowledge Building community was forming. Furthermore, we observed early attempts at collective theory building. In one thread discussing how salmon return to their birthplace, students built on each other:

Student A:	My theory - salmon can smell
Student B:	My theory - They memorize the smell of the place they were born in
Student C:	They can all lay 3000 eggs in the same place and there will be 3000 *so it will be 3000 x 3000 and then hunman can get the really big salmon*

This indicated that while reproduction was still high, the community was becoming "safe" enough for risk-taking, a prerequisite for the "design mode."

 © ISLS

Study 3: Cross-community co-creation

Study 3 sought to deepen agency by creating a networked public. The initial design (Oct-Dec 2021) kept the Grade 2 and Grade 9 classes in separate Knowledge Forum views to prevent disruption. However, this separation created an "othering" effect. The Grade 9 students viewed the Grade 2s as subjects to be observed rather than intellectual partners, while the Grade 2s lacked access to the sophisticated vocabulary and technical knowledge of the older students. Ideas generated in one view effectively "died" in the silo, with no mechanism for uptake or improvement by the other group. The lack of a shared workspace meant that epistemic agency was restricted to the local level, failing to leverage the distributed expertise of the cross-age partnership. Over four months, note production was low (126 notes for Grade 9s).

In Term 2, a significant design change moved both classes into a single, shared Knowledge Forum view. The impact was immediate: the unified community created 595 notes in four months. The classes connected with domain experts—Grade 9s with a biologist on sustainable cities, and Grade 2s with an expert on indigenous salmon stories. This set the stage for a culminating design challenge: "How would you design a new neighborhood?"

The co-design process began with collaborative brainstorming in the shared space, which was then translated by the Grade 9 students into a shared Minecraft simulation (Figure 1).

Figure 1

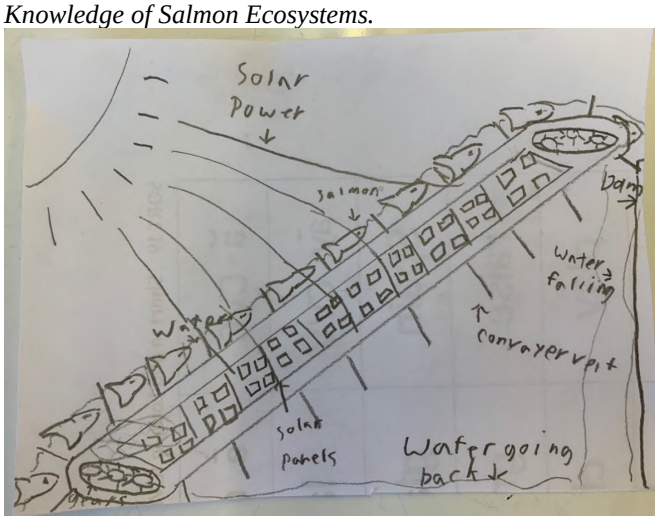
The Translation of Community Ideas into a Shared Artifact: The Grade 9 Students' Minecraft Simulation of the Co-Designed Neighborhood.



This partnership facilitated a shift to "design mode," enabling deep knowledge synthesis. The Grade 9 students confirmed this in interviews, noting that the Grade 2s "came up with the ideas, we researched and tried to see how it could work." This bond led to rotating leadership (Ma, Matsuzawa, & Scardamalia, 2016); at times Grade 2s led on salmon ecology, while Grade 9s led on energy sustainability. A powerful example of this synthesis is a Grade 2 student's design for a "solar-powered elevator to circumvent human dams" (Figure 2).

Figure 2

An Example of Knowledge Synthesis: A Grade 2 Student's Design for a "Solar Powered Elevator," Integrating Grade 9 Concepts of Sustainable Energy with Grade 2 Knowledge of Salmon Ecosystems.



 © ISLS

This artifact represents a "rise-above" synthesis. The student explicitly integrates the Grade 9 students' research (solar power) with their own deep contextual knowledge (salmon migration). Unlike in Studies 1 and 2, the student is not repeating facts but proposing a novel technical solution, demonstrating a clear emergence of epistemic agency.

Discussion and implications

This study demonstrates that cultivating epistemic agency is a progressive, intentionally designed process.

1. *The Power of the Public Community:* A key implication is that agency is fostered when the "public" for student ideas expands. Students moved toward sustained creative work when they understood their ideas served a real partner (the Grade 9 class). This shifts the relationship between students and learning from consumption to collective responsibility. Fundamentally, this shifts the relationship between students and learning as they develop greater ownership and willingness to self-direct their inquiry.
2. *Multimodality as an Equity Bridge:* Multimodality served as a critical tool for equity. In Study 3, drawings and simulations acted as a common language, allowing Grade 2 and Grade 9 students to collaborate meaningfully despite age differences, validating diverse forms of literacy. This dynamic disrupted the traditional "expert-novice" hierarchy often seen in cross-age tutoring. Typically, the older student "teaches" the younger. Here, a reciprocal interdependence emerged: the Grade 2s provided the creative constraints based on their deep empathy for the salmon, while the Grade 9s provided the technical solutions based on their research into sustainability. The shared visual language of Minecraft and drawing leveled the playing field, allowing the younger students to lead the conceptual design while the older students led the technical implementation.
3. *Analytical Tools for Knowledge Building:* Finally, the StAS tool offered a diagnostic capability, identifying the prevalence of information reproduction in early phases. Future work should focus on developing such analytics to track the "rise-above" quality of synthesis seen in Study 3, further bridging the gap between classroom practice and Knowledge Building theory.

Limitations and future work

While this longitudinal study provides a rich account of cultivating epistemic agency, the findings should be interpreted within two specific boundaries.

First, the context was highly favorable: a single independent school with small class sizes and a dedicated teacher who participated across all three years. While this continuity allowed for deep iteration, it limits the immediate generalizability of the findings to resource-constrained public settings. Future work should explore how these design principles—specifically the cross-age partnership—scale in larger, more diverse classrooms where coordinating schedules between Grade 2 and Grade 9 poses greater logistical challenges.

Second, the StAS tool, while promising, serves as a proxy for epistemic change rather than a direct measure of idea quality. The tool measures semantic distance from authoritative texts; it does not inherently distinguish between a "misconception" (which is also semantically distant) and a "rise-above" synthesis. In this study, qualitative coding was required to validate that lower StAS scores corresponded to productive idea generation. Future development of the tool aims to integrate "promisingness" filters (Chen et al., 2015) to automatically distinguish between noise and genuine knowledge advancement.

Finally, regarding the qualitative analysis, we acknowledge the limitation of a single coder. To mitigate bias, we employed prolonged engagement in the field and peer debriefing strategies. Future iterations of this work will employ multiple coders to establish inter-rater reliability, particularly when classifying specific types of epistemic moves in the discourse.

References

Bereiter, C., & Scardamalia, M. (2003). Surpassing ourselves: An inquiry into the nature and implications of expertise. Open Court.

Brown, A. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. The Journal of Learning Sciences, 2(2), 141–178.

Charmaz, K. (2006). Constructing grounded theory: A practical guide through qualitative analysis. Sage Publications.

Chen, B., Scardamalia, M., & Bereiter, C. (2015). Advancing knowledge-building discourse through judgments of promising ideas. International Journal of Computer-Supported Collaborative Learning, 10(1), 23–44.

Gloor, P. (2006). Swarm Creativity: Competitive Advantage Through Collaborative Innovation Networks. Innovation, 8(4). https://doi.org/10.5172/impp.2006.8.4.413

Herrington, J., & Kervin, L. (2007). Authentic learning supported by technology: Ten suggestions and cases of integration in classrooms. Educational Media International, 44(3), 219-236.

Immordino-Yang, M. H. (2016). Emotion, sociality, and the brain's default mode network: Insights for educational practice and policy. Policy Insights from the Behavioral and Brain Sciences, 3(2), 211-219.

Khanlari, A., Huang, R., Scardamalia, M. (2024). Toward AI-Driven Knowledge Building Analytics to Assess Idea Paths and Potential for Idea Transformation. In Proceedings of the Knowledge Building Summer Institute @ Singapore 2024. Online: Knowledge Building International.

Lewis, S., Pea, R., & Rosen, J. (2010). Beyond participation to co-creation of meaning: mobile social media in generative learning communities. Social Science Information, 49(3), 351-369.

Ma, L., Matsuzawa, Y., & Scardamalia, M. (2016). Rotating leadership and collective responsibility in a grade 4 Knowledge Building classroom. International Journal of Organisational Design and Engineering, 4(1-2), 54-84. https://doi.org/10.1504/IJODE.2016.080159

Scardamalia, M., & Bereiter, C. (2003). Knowledge Building. In Encyclopedia of Education (2nd ed., pp. 1370-1373). Macmillan Reference USA.

Scardamalia, M. (2002). Collective cognitive responsibility for the advancement of knowledge. Liberal Education, 88(4), 60-66.

Scardamalia, M. (2004). CSILE / Knowledge Forum. In Education and Technology: An Encyclopedia (pp. 183–192). ABC-CLIO.

Scardamalia, M., & Bereiter, C. (2006). Knowledge Building: Theory, Pedagogy, and Technology. In R. K. Sawyer (Ed.), The Cambridge Handbook of the Learning Sciences (pp. 97–118). Cambridge University Press. https://doi.org/10.1598/RT.61.2.5

Scardamalia, M., & Bereiter, C. (2022). Knowledge Building and Knowledge Creation. In R. K. Sawyer (Ed.), The Cambridge Handbook of the Learning Sciences (2nd ed., pp. 385–405). Cambridge University Press.

International Society of the Learning Sciences

Reconceptualize Conversational AI as a Learning Medium, From a Holistic Learning Perspective

Yuqi Wang, University of Toronto, yuqiw.wang@mail.utoronto.ca

Abstract: Current university policies predominantly restrict conversational AI use in coursework, based on assumptions that students engage with AI under transmission orientations. However, these policies overlook transactional pedagogical possibilities. This study reconceptualizes conversational AI as a medium for holistic learning through structured transactional engagement rather than information delivery. Drawing upon Miller's Holistic Learning paradigm and Computer-Supported Collaborative Learning (CSCL) principles, we conducted semi-structured interviews with 12 educators enrolled in a Holistic Learning postgraduate course to examine: (1) current perceptions and concerns, (2) capacity to distinguish transmission from transactional AI applications, and (3) necessary implementation conditions. Findings demonstrate most educators readily recognized conversational AI's pedagogical function depends on prompt design rather than technological properties. I propose integrating CSCL script principles with Miller's six connection types to operationalize transactional AI through multi-level prompt design. This framework moves beyond the ban-or-permit binaries toward nuanced, pedagogically-grounded implementation that honors both academic integrity and innovative learning design.

Introduction

Students increasingly use conversational AI tools—Natural Language Processing systems designed for sustained, multi-turn dialogue interactions (e.g., ChatGPT, Claude, Gemini)—for both everyday inquiry and academic coursework, yet universities face divergent responses to this practice. Most institutions delegate restriction authority to individual instructors, allowing course-level bans on conversational AI use based on concerns about academic integrity, information accuracy, and passive learning dependency (Loyola University New Orleans, 2024; Northeastern University, 2025; University of New South Wales, 2025). Others, such as Stanford Graduate School of Business, explicitly prohibit instructors from banning conversational AI in take-home coursework (Stanford Graduate School of Business, 2025). These policy debates, however, rest on an implicit assumption: conversational AI use primarily involves transmission-based learning in which students passively receive AI-generated content to substitute for their own thinking. This framing positions conversational AI as either a productivity tool (to be permitted with attribution) or a cheating mechanism (to be prohibited). Yet this binary overlooks alternative pedagogical possibilities. What if conversational AI could function not as an information source, but as a medium for structured transactional engagement to support learner-driven dialogue, connection-building, and inquiry across multiple knowledge domains?

This paper draws on Miller's holistic learning paradigm (Miller, 2019), a theoretical framework largely absent from current educational technology debates, to reconceptualize conversational AI's role in learning. Miller distinguishes between three curricular orientations: transmission (transfer of information), transaction (learning through dialogue and critical thinking), and transformation (deep personal and social change through interconnection). While current university policies appropriately restrict transmission-based Conversational AI use, they fail to distinguish this from transactional applications where conversational AI serves as a collaborative learning partner facilitating meaningful connections across knowledge domains. Despite growing policy debates, no empirical research has examined educators' perspectives on conversational AI through holistic learning frameworks. We address this gap through semi-structured interviews with 12 educators enrolled in a holistic learning postgraduate course. This gap carries particular consequence for CSCL research: without frameworks distinguishing transmission from transactional AI engagement, designers lack principled guidance for scaffolding human-AI dialogue toward transactivity, and the field lacks empirical grounding for moving beyond binary policy responses. Our study examine three research questions:

(RQ1) How do educators perceive conversational AI's current role and associated concerns?

(RQ2) Can educators distinguish between transmission-based and transactional applications of conversational AI when introduced to Miller's holistic learning framework?

(RQ3) What conditions do educators identify as necessary for conversational AI to support transactional learning rather than transmission-based information retrieval?

 © ISLS

Theoretical frameworks

Holistic learning orientations: Transmission, transaction and transformation

Miller's holistic learning framework (Miller, 2019) provides a systematic theoretical foundation for reconceptualizing conversational AI's pedagogical function beyond transmission-based applications. The three metaorientations of holistic learning - transmission, transaction and transformation - offers a developmental framework for understanding learning progression. Transmission involves "one way of imparting information to the student by the teacher or the text," while the transaction is "more interactive" and "involves dialogue between teacher and student or student and student" with focus on "problem solving and the development of cognitive skills" (Miller, 2019). This orientation also resonates with Deweyan views of inquiry, interaction, and learning through experience (Dewey, 1938).Transformation represents the most sophisticated level, involving "connecting to and nurturing the inner, imaginative life of the student" where "the knower and known become one; what we study becomes part of our being and does not remain separate" (Miller, 2019). Miller operationalizes these orientations through six educational connections: subject, community, earth, thinking, body-mind, and soul-heart connections. Miller emphasizes that "the holistic teacher will use all three of these approaches to develop a rhythm in their teaching" (Miller, 2019). Current AI prohibition policies address transmission-based concerns but fail to recognize transactional possibilities.

Conversational AI as a Learning medium, not a tool

Understanding conversational AI's transactional role requires distinguishing tools from mediums. McLuhan (1964) established that media are not neutral conduits but active shapers of perception and cognition—the medium itself fundamentally alters both message and recipient regardless of content transmitted. Tools align with transmission—students employ them instrumentally to achieve predetermined outcomes, terminating engagement upon goal completion. Mediums align with transaction—interaction generates emergent learning extending beyond initial objectives. Applying this distinction, Ch'ng (2024) illustrates that treating conversational AI as a tool is inadequate, as "When you use [conversational] AI, in reading the outputs, your mind has already been changed. You have consumed the output of AI" (Ch'ng, 2024). The medium shapes thinking patterns and knowledge engagement modes regardless of information accuracy. This medium-based perspective aligns with the computers as social actors (CASA) paradigm (Nass et al., 1994), demonstrating that individuals unconsciously apply social interaction rules to communicative systems.

Structuring human-AI collaboration for transactional learning

Recognizing conversational AI as medium rather than tool necessitates examining how human-AI relationships preserve learner agency while leveraging dialogic capacities. Lodge et al. (2024) map these relationships across cognitive function (offloading to extending capabilities) and interaction mode (individual to collaborative), identifying four configurations: cognitive offloading, extended mind, co-regulation, and hybrid learning. For transactional learning, the collaborative-extending quadrant enables conversational AI to participate in dialogue extending thinking while maintaining transparency about its interpretive role. Yan (2025) specifies four escalating agency levels: adaptive instruments, proactive assistants, co-learners, and peer collaborators. Higher agency levels—co-learner and peer collaborator—prove most conducive to transactional learning through sustained dialogue and perspective-taking. Computer-supported collaborative learning (CSCL) research provides critical frameworks for operationalizing dialogue quality in these collaborative configurations. Central to CSCL is the concept of transactivity, defined as "building on learning partner's contributions, and engaging in and maintaining coherent and joint reasoning" (Hesse, 1997). Transactive engagement involves learners actively questioning, elaborating, and integrating contributions from learning partners rather than passively receiving information. Weinberger and Fischer's (2006) collaboration scripts demonstrate how structured interaction protocols can promote transactivity through specific design principles: regulation of learning activities, process-oriented instruction, and awareness induction. Despite theoretical frameworks suggesting conversational AI's capacity for transactional learning, empirical research examining how educators conceptualize AI through holistic learning orientations remains blank.

Methods

We conducted semi-structured interviews with 12 educators enrolled in a postgraduate course on Holistic Learning: seven practicing middle and high school teachers in Ontario, two doctoral students in Curriculum and Pedagogy, two student teachers, and one university instructor, representing diverse teaching contexts and career stages. Interviews were conducted in-person, and were audio-recorded and transcribed verbatim. Interviews

followed a two-phase structure. Phase one elicited spontaneous perspectives through open-ended questions about AI experiences and concerns (RQ1). Phase two introduced Miller's transmission-transaction-transformation model only after Phase one concluded, presenting participants with a brief written summary before inviting assessment of AI's potential across his six connection types (RQ2, RQ3), minimizing theoretical priming of earlier responses. We employed reflexive thematic analysis (Braun & Clarke, 2006) using a hybrid approach: Phase one data were analyzed inductively to surface educator concerns, while Phase two data were interpreted deductively against Miller's orientations and CSCL constructs. Themes were stabilized through iterative reading, memoing, and peer debriefing.

Results and discussion

RQ1: Educators' perceptions of conversational AI's current role and associated concerns

Educators described an educational landscape where conversational AI remains largely absent from pedagogical practice despite increasing availability. As one secondary teacher noted, while "there are some AI tools in the schools already" for administrative or teacher-facing purposes, such as tools that "help us [teachers] to give formative feedback". But the reality is that "pedagogical use of AI, I don't see many in conversational AI yet" (P1). This absence reflects both technological and pedagogical factors: schools lack systematic frameworks for implementation, and educators harbor legitimate concerns about student usage patterns.

When asked about these concerns, educators articulated two interconnected apprehensions that align with transmission-based usage scenarios. First, information integrity concerns emerged prominently. One teacher captured widespread unease: "I agree some teachers are fearful of chatbots. Like presenting cheat sheets of all answers to the students" (P2). This fear reflects documented issues with AI hallucination rates and training data biases. When students approach AI as an authoritative information source—what we characterize as tool-based usage—educators cannot reasonably expect students to possess domain expertise necessary for systematic fact-checking. The concern is not AI's existence but its positioning as answer provider rather than dialogue partner. Second, educators expressed deep concern about learning process authenticity. Multiple participants emphasized that acquiring answers circumvents essential cognitive development. As one educator posed the question: "What matters more: the answers or the thinking process?" (P2). Another made this concern explicit: "AI can show you the results. But what is the process?" (P8). These questions reflect understanding that transmission-based AI usage—where students seek definitive responses—compromises "tolerance for uncertainty and independent reasoning development" that constitute fundamental learning mechanisms. One teacher articulated this concern in terms of internalization and assessment validity: "If students just jot down the answer without internalizing it, the assessments mean nothing" (P2). Another emphasized student agency: "AI can do a lot of things, but people need to learn it by themselves. AI can't learn for the students" (P7). These perspectives reveal sophisticated understanding that the problem is not AI providing information, but students treating AI as a tool for bypassing intellectual struggle. Critically, these concerns are empirically grounded and pedagogically legitimate.

RQ2: Distinguishing transmission-based and transactional applications through a holistic learning perspective

When invited to consider AI through holistic learning frameworks, educators spontaneously distinguished between product-focused and process-focused engagement. The educator who posed "What matters more: the answers or the thinking process?" elaborated: "We are learning it for today, for this moment, for presence, for this curiosity" (P2), suggesting a shift from answer acquisition to inquiry-centered engagement. This same educator recognized corresponding assessment implications: "The problem is the assessment format is wrong. If students just jot down the answer without internalizing it, the assessments mean nothing" (P2). Most significantly, when explicitly asked to map conversational AI onto Miller's three holistic learning orientations, one teacher stated: "Chatbot can do transmission, transactions and transformation. It just depends on how students respond to it" (P2, emphasis added). This educator elaborated transactional possibilities: "Internalization and inquiry-based learning. It is adaptive" (P2). Another emphasized sustained exploration: "Conversational AI can be quite powerful in chatting with students, exploring some topics deeply" (P5). These responses indicate recognition that AI's pedagogical function depends on prompt design and student engagement patterns rather than inherent technological properties.

RQ3: Necessary implementation conditions for supporting transactional learning

Participants recognized Miller's connection types as operational pathways for transactional AI implementation, extending beyond purely cognitive engagement. One educator explicitly noted: "Social-emotional part, ChatGPT can be great" (P3), challenging assumptions limiting AI to information processing and suggesting affective dimensions. Educators implicitly mapped AI's transactional potential across six connection domains. Subject connections enable cross-disciplinary "exploration following student inquiry" (P11) rather than predetermined sequences. Community connections provide "low-stakes" (P4) dialogue practice addressing "power dynamics" (P4) and judgment concerns inhibiting transactivity with human educators. Thinking connections leverage AI's capacity to introduce alternative perspectives, contingent upon students actively interrogating "rather than passively accepting information" (P5). For body-mind and soul-heart connections, participants acknowledged AI's capacity for prompting reflection on "embodied experiences and personal meaning-making" (P7) while recognizing fundamental limitations in engaging existential dimensions requiring consciousness. Participants emphasized prompt design as a critical implementation factor. As noted above, AI's pedagogical function "depends on how students respond to it" (P2), with distinction between "exploring some topics deeply" (P5) versus answer-seeking queries. This suggests transactional implementation requires explicit instruction in engagement strategies fostering sustained dialogue rather than single-turn information retrieval.

Discussion

Educators' recognition that AI's pedagogical function "depends on how students respond to it" (P2) reveals critical insight: transactional versus transmission-based usage emerges not from technological properties but from prompt design. This finding aligns with our theoretical framework positioning conversational AI as medium rather than tool (Ch'ng, 2024). When engagement terminates upon receiving information, AI functions as a tool aligned with transmission orientation. When interaction generates emergent learning through sustained dialogue, AI functions as a medium aligned with transactional orientation.

The process-product distinction educators articulated—learning "for this moment, for this curiosity" (P2) versus answer acquisition—directly maps onto Miller's (2019) transmission-transaction differentiation. This reframing transforms the concerns identified in RQ1: when pedagogical objectives center on inquiry processes rather than information acquisition, integrity concerns become learning opportunities (AI errors as chances for critical evaluation) and authenticity concerns shift from independent answer generation to genuine engagement in sustained inquiry. Drawing on the computers as social actors paradigm (Nass et al., 1994), this positions conversational AI as collaborative learning partner rather than information source. Educators implicitly recognized the need for structured prompt design of conversational AI, noting students must avoid "passively accepting information" (P5) and that AI's value depends on "how students respond" (P2). However, participants' acknowledged implementation remains nascent: "it is still at its early stage of discovery" (P10).

I propose integrating CSCL collaboration script principles (Kobbe et al., 2007; Weinberger, 2011; Weinberger & Fischer, 2006; Vogel et al., 2017) with holistic learning orientations to operationalize transactional AI through multi-level prompt design. Vogel et al. (2017) suggested scene-level scripts may effectively support domain-specific knowledge acquisition, while scriptlet-level scripts foster collaboration skills. Applied to conversational AI contexts, play-level design establishes overall dialogic stance positioning AI as collaborative partner rather than answer source. Scene-level prompts structure phase-specific inquiry aligned with Miller's (2019) connection types: exploration phases for subject connections (cross-disciplinary inquiry), elaboration phases for thinking connections (perspective examination), and synthesis phases for community connections (dialogue integration). Scriptlet-level prompts scaffold specific cognitive operations: transactive questioning ("You suggested X. However, Y. How would you reconcile these?"), justification requests ("What evidence supports this claim?"), and metacognitive monitoring ("How might your framing influence my response?"), supporting body-mind and soul-heart connections through reflective engagement. This integration of CSCL scripting principles into conversational AI prompt design enables systematic instruction tailored to pedagogical objectives, transforming AI from answer source to dialogue partner through structured intervention consistent with Lodge et al.'s (2024) collaborative-extending configuration.

Educational institutions face urgent policy decisions about conversational AI amid rapid technological change. This research suggests that prohibition debates rest on incomplete pedagogical frameworks. The CSCL-informed prompt design framework I proposed operationalizes Miller's (2019) holistic learning orientations in conversational AI contexts, supporting transactional learning while restricting transmission-based misuse. This approach enables cultivation of the six connection types—subject, community, thinking, body-mind, earth, and soul-heart—through which holistic learning is realized. By reconceptualizing conversational AI through holistic learning orientations and operationalizing transactional applications through CSCL collaboration scripts, we can move beyond the binary of ban-or-permit toward nuanced, pedagogically-grounded implementation that honors both academic integrity and innovative learning design.

References

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). MIT Press.

Ch'ng, E. (2024, January 10). *Is AI a tool or a medium?* [Video]. YouTube. https://www.youtube.com/watch?v=anezucvMOzo

Dewey, J. (1938). *Experience and education*. Macmillan.

Hesse, F. W., Garsoffky, B., & Hron, A. (1997). Interface-design für computerunterstütztes kooperatives lernen. In L. J. Issing & P. Klimsa (Eds.), *Information und lernen mit multimedia* (2nd ed., pp. 253–267). Beltz.

Kobbe, L., Weinberger, A., Dillenbourg, P., Harrer, A., Hämäläinen, R., Häkkinen, P., & Fischer, F. (2007). Specifying computer-supported collaboration scripts. *International Journal of Computer-Supported Collaborative Learning, 2*(2-3), 211-224. https://doi.org/10.1007/s11412-007-9014-4

Lodge, J. M., Yang, S., Furze, L., & Dawson, P. (2023). It's not like a calculator, so what is the relationship between learners and generative artificial intelligence? *Learning: Research and Practice, 9*(2), 117–124. https://doi.org/10.1080/23735082.2023.2261106

Loyola University New Orleans. (2024). *Artificial intelligence policy & guidance*. https://academicaffairs.loyno.edu/default/files

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 72-78). ACM. https://doi.org/10.1145/191666.191703

Northeastern University. (2025). *Generative AI: Academic integrity and course policy guidance*. https://provost.northeastern.edu/ai-guidance

McLuhan, M. (1964). *Understanding media: The extensions of man.* McGraw-Hill.

Miller, J. P. (2019). *The holistic curriculum* (3rd ed.). University of Toronto Press.

Northeastern University. (2025). *Policy on the use of artificial intelligence systems* (Policy No. 125). https://policies.northeastern.edu/policy125/

Stanford Graduate School of Business. (2025). *Course policies on generative AI use*. Stanford University Teaching and Learning Hub. https://gsb-sites.stanford.edu/teaching-learning-hub/teaching-and-ai/course-policies-generative-ai-use

Stegmann, K., Weinberger, A., & Fischer, F. (2007). Facilitating argumentative knowledge construction with computer-supported collaboration scripts. *International Journal of Computer-Supported Collaborative Learning, 2*(4), 421–447.

University of New South Wales. (2025). *Using AI in university*. https://www.unsw.edu.au/student/managing-your-studies/academic-skills-support/toolkit/ai

Vogel, F., Wecker, C., Kollar, I., & Fischer, F. (2017). Socio-cognitive scaffolding with computer-supported collaboration scripts: A meta-analysis. *Educational Psychology Review, 29*(3), 477-511. https://doi.org/10.1007/s10648-016-9361-7

Weinberger, A. (2011). Principles of transactive computer-supported collaboration scripts. *Nordic Journal of Digital Literacy, 6*(3), 189–202. https://doi.org/10.18261/ISSN1891-943X-2011-03-06

Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & Education, 46*, 71–95.

Yan, L. (2025). From passive tool to socio-cognitive teammate: A conceptual framework for agentic AI in human-AI collaborative learning. *arXiv*. https://arxiv.org/abs/2508.14825

International Society of the Learning Sciences

Scoping Review of Knowledge-Creation Practices Supported by AI: Identification of Research Strands of Different Communities

Jun Oshima, RECLS, Shizuoka University, joshima@inf.shizuoka.ac.jp
Shotaro Naganuma, Kyushu University, naganuma.shotaro.062@m.kyushu-u.ac.jp

Abstract: This scoping review examines how AI supports collaborative learning as knowledge creation in varied educational contexts. Unlike previous reviews that discuss learning more broadly, this study focuses on the epistemological and methodological foundations of AI-supported collaboration. Using the PRISMA protocol, the authors analyzed 72 empirical studies published between 2020 and 2025, employing bibliographic coupling and ENA to identify research communities and their research strands. Six research communities emerged, each representing different strands—ranging from cognitive and sociocultural to design-based and systemic approaches. These communities reveal diverse perspectives on AI's role. The findings indicate that much of the current research centers on individual learning and distal outcomes, while studies oriented toward knowledge creation, especially within CSCL and the LS, remain limited but are increasing. The review highlights the importance of future research prioritizing multimodal, group-level AI collaboration designs, grounded in robust theoretical and methodological foundations, to foster knowledge creation practices.

Introduction

In this scoping review, we aimed at exploring AI-supported collaboration with two specific perspectives: (1) the focus on collaborative learning as knowledge creation, and (2) research traditions. Systematic reviews so far have discussed learning in the generic form and have not examined the epistemological perspectives that reviewed studies hold. From the learning sciences perspective, in particular, the knowledge-creation metaphor of collaborative learning has been developed as a new goal of instruction and has led to the development of CSCL systems. It frames learning as a collaborative process of producing and improving knowledge valuable to a community (Paavola & Hakkarainen, 2005) and emphasizes inquiry, innovation, and collective responsibility for ideas, viewing knowledge as dynamic and evolving through interaction and co-construction rather than static content to be acquired or practiced (Scardamalia & Bereiter, 2022). This view positions learners as active knowledge builders, engaging in creative and problem-solving processes similar to those found in scientific and professional communities. A further systematic review of studies focused on knowledge-creation practices can provide new directions for collaborative learning supported by AIs, aligning with the goal of instruction in the 21st century.

Besides the specific attention to the epistemological perspective, research traditions (Enyedy & Stevens, 2022) in studies are essential to rigorously provide scientific evidence for a new style of collaborative learning supported by AIs and to feed insights back into the future design of learning environments. The *collaboration-as-a-window-onto-thinking* tradition reflects a cognitivist epistemology, viewing collaboration as a way to access internal mental processes. Here, dialogue acts much like a "think-aloud" protocol, making interaction largely incidental. The *collaboration-for-distal-outcomes* tradition adopts an effect-oriented epistemology, investigating how certain discourse patterns (e.g., IRE, exploratory, or accountable talk) are linked to individual learning outcomes such as test performance or motivation. The *collaboration-for-proximal-outcomes* tradition draws on sociocultural and interactionist epistemologies, focusing on how learning emerges through the interaction itself via intersubjectivity, tool appropriation, or the use of shared referents (Roschelle, 1992; Enyedy, 2003). Finally, the essential methodological approach named as *collaboration-as-learning* is to rigorously gather scientific evidence of a new style of collaborative learning supported by AIs and feed insights back into the future design of tradition is rooted in distributed and ethnomethodological epistemologies, treating collaboration as both the unit and outcome of learning, and emphasizing the internal organization of activity systems and collective accomplishments (Barron, 2003; Scardamalia & Bereiter, 2022). Each tradition offers a unique perspective on collaboration, shaped by its epistemological stance, and influences what is considered evidence of learning.

Methodologically, we not only followed the standard PRISMA protocol but also examined existing research strands in target studies using Epistemic Network Analysis (ENA) (Shaffer et al., 2016). Nachtigall et al. (2022) employed ENA to capture the relationships between design features (e.g., collaboration, inquiry-based tasks, real-world contexts) and the intentions of authenticity they served (e.g., emulating professional practices,

 © ISLS

connecting with communities of practice). This allowed them to visualize and compare patterns of co-occurrence of codes across studies with higher versus lower effectiveness, revealing how particular combinations of design elements were linked to cognitive and motivational outcomes. Their use of ENA demonstrates how network-based methods can enrich literature reviews by revealing systemic interactions that would remain invisible in traditional frequency-based analyses. We further implemented a new algorithm for the ENA modeling process, specifically the conversation. The approach by Nachtigall et al. (2022), treating each study as a "conversation" in their modeling, could not actually capture a conversation as a dialogue unit among studies. Through bibliographic coupling network analysis, we identified research communities based on shared references (Hmelo-Silver & Jeong, 2021). These communities served as datasets for calculating code co-occurrences, effectively modeling conversations in ENA. Combining bibliographic coupling analysis with ENA enabled visualization of how different research communities align with particular strands defined by specific epistemological perspectives and methodological traditions.

In summary, the purpose of our scoping review is to examine current research strands in studies of AI-supported collaborative learning and to propose new directions for future research. The following research questions guide this investigation:

RQ1: *How are studies on collaborative learning as knowledge creation, supported by AIs, grouped into research communities?*

RQ2: *How do the research communities differ in their research strands?*

Method

Article identification through the PRISMA protocol

Our search strategy consisted of two groups of terms. The first group targeted various forms of collaborative learning, referring to Mende et al. (2021), while the second group limited the results to studies involving artificial intelligence. The final search string was:

Topic = (collaborati OR cooperati* OR group OR peer OR dyad) AND "artificial intelligence"

This search yielded 29,208 results. After limiting to English-language publications in Q1-ranked journals in the two categories ("EDUCATION & EDUCATIONAL RESEARCH" and "COMPUTER SCIENCE, INTERDISCIPLINARY APPLICATIONS), 1,298 articles remained.

The screening process was carried out in two stages—abstract and full-text screening. At the abstract stage, 10% of the records (130 articles) were independently screened by two reviewers, achieving a Cohen's kappa of .714. Discrepancies were resolved through discussion, and the remaining records were divided for single screening. This process resulted in 441 articles meeting the criteria in Table 1. At the full-text screening stage, 10% of the articles (44) were double-screened, yielding a Cohen's kappa of .490. Disagreements were resolved by consensus, and the remaining texts were divided equally between the two reviewers. After applying an additional inclusion criterion requiring learner engagement in knowledge-creation processes (e.g., creation and development of material or conceptual artifacts, or conscious knowledge advancement, discovery, and innovation), 72 articles were retained for coding.

Coding of articles for ENA

After identifying research communities through bibliographic coupling analysis (Waltman et al., 2010) using *VOSviewer* (https://www.vosviewer.com/), we coded the selected articles using four dimensions: (1) collaboration context, (2) AI's role, (3) methodological traditions, and (4) analytical modality of collaboration. We first jointly coded a sample of ten articles and used a social moderation procedure (Frederiksen et al., 1998) to ensure complete agreement on all codes. The remaining articles were then independently coded by the first and second authors.

The collaboration context concerns the type of collaboration involving AI. A single learner can use AI for individual learning—a new collaborative possibility enabled by AI, especially generative AI. We refer to this as "Human-AI" collaboration. Alternatively, "Group-AI" collaboration involves AI in activities where multiple learners work together. The AI's role examines how learners use AI for collaborative knowledge creation. We developed six codes based on the categories suggested by Sharples (2023): Possibility Engine, Socratic Opponent, Collaboration Coach, Co-designer, Exploratorium, and Storyteller. Since Sharples (2023) does not provide a detailed explanation for Collaboration Coach, we referred to Mikroyannidis and Armakolas (2025), which cites

Sharples to generate a similar code. We chose to define Collaboration Coach as an AI use focused on regulating collaboration, rather than generating idea-level epistemic output like Possibility Engine or Socratic Opponent. The methodological traditions address the methodological approaches used to analyze both "Human-AI" and "Group-AI" collaboration. We referenced Enyedy and Stevens (2022), who identified four methodological traditions to studying collaborative learning: collaboration-as-a-window, collaboration-for-distal-outcomes, collaboration-for-proximal-outcomes, and collaboration-as-learning. Finally, we referred to the "Definitional Framework of Modalities in Multimodal Collaboration Analytics Literature" by Esterhazy et al. (2025). To capture various types of behavior during collaborative activities, they conducted a scoping review and classified the modalities used in multimodal collaboration analytics into five categories: Auditory, Visual, Kinesthetic, Tactile, and Physiological.

Results and discussion

Identification of research communities: Bibliographic coupling analysis

A bibliographic coupling analysis was conducted to cluster the target articles based on shared references. We identified seven communities; however, because one community contained only a single article (n=1), we excluded it from subsequent analyses.

Studies cited in the six communities present distinct yet complementary perspectives on the evolving role of artificial intelligence (AI) in education. Articles cited in Community 1 (n=16) focus on the micro-level of interactional learning, positioning ChatGPT as a dialogic partner that supports the co-construction of knowledge and reflective feedback, with an emphasis on social presence and authenticity in AI-mediated writing. In contrast, those in Community 2 (n=14) adopt a macro-level, policy-oriented approach, exploring AI's pedagogical potential alongside ethical, regulatory, and epistemic risks. Articles cited in Community 3 (n=13) bridge these viewpoints, examining feedback ecosystems and collaborative learning where AI analytics enhance metacognitive awareness and shared inquiry. Articles cited in Community 4 (n= 12) shift to an applied, experiential perspective, integrating AI into game-based environments to optimize motivation, flow, and cognitive support through adaptive feedback.

Articles cited in Community 5 (n=10) highlight innovation and design thinking, portraying AI as a co-designer blending analytical precision with human creativity and empathy to solve complex educational and organizational problems. Finally, articles cited in Community 6 (n=6) stand out for their synthetic and integrative approach, drawing on diverse empirical studies to depict AI as a comprehensive educational ecosystem that fosters inclusivity, creativity, and self-efficacy across disciplines. While all the cited articles view AI as transformative, they differ in scale—from individual cognition and classroom interaction to institutional governance and systemic innovation—and in emphasis, ranging from practical pedagogy to ethical reflection and creative co-design. Collectively, these distinctions illustrate the multifaceted nature of AI in education: a feedback mechanism, a learning collaborator, a design tool, and a catalyst for broader social and ethical transformation.

Research strands of identified communities: Epistemic network analysis

We defined each unit of analysis as all data lines associated with a single research community value, subsetted by ID. The resulting networks for each unit (i.e., each article in this case) were aggregated within the model. Network aggregation was performed using binary summation, indicating the presence or absence of co-occurrence between each pair of codes per line. The ENA model then normalized the networks for all units before applying dimensional reduction to account for varying numbers of coded lines across units. Dimensional reduction was achieved using singular value decomposition, yielding orthogonal dimensions that maximized the explained variance (17.5% for SVD1, 12.1% for SVD2). Our model achieved co-registration correlations of 0.95 (Pearson) and 0.96 (Spearman) for SVD1, and 0.93 (Pearson) and 0.94 (Spearman) for SVD2, indicating a strong fit between the visualization and the original model.

To test for differences between the mean networks of communities, we conducted multiple comparisons using Mann-Whitney tests on the location of the centrodes of research communities in the projected ENA space. All clusters were significantly different from one another in either the first or the second dimension. It suggests that all research communities have their unique research strands.

To explore differences in research strands among the communities, we visualized two comparison networks in the ENA space. The comparison networks illustrate how connections between codes differ across communities, highlighting distinct features of each group. Figure 1 (left) shows a comparison between community 3 and community 6, which are the most distant along the first dimension. This dimension represents "research traditions" and "collaboration contexts." The positive end of the first dimension is primarily characterized by codes such as *collaboration for proximal outcomes*, *collaboration as learning*, and *Group-AI*. The negative end is marked by *Human-AI*, with no specific research tradition codes. Figure 1 (right) presents a comparison between

community 1 and community 4, which are maximally separated along the second dimension. This dimension reflects "AI's role" and "analytical modality of collaboration." Here, the positive side is defined by *Co-designer* and *Collaboration coach*, while the opposing side is associated with *Possibility engine* and *Socratic Opponent*, along with the analytical modality, *Tactile*.

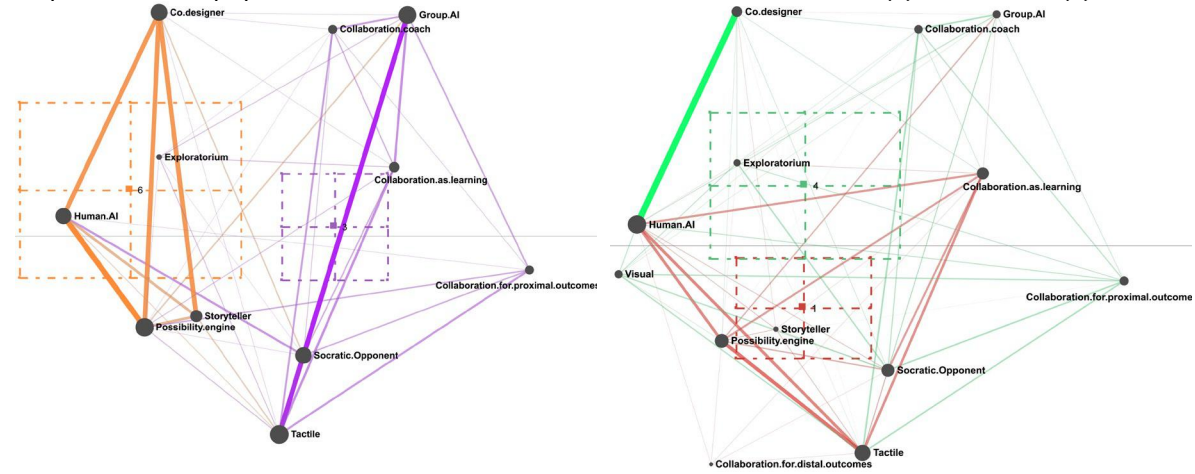
Thus, different communities have developed as unique research strands. A few strands (clusters 4, 5, and 6) were characterized as cohorts of studies focusing on the implementation of AIs without rigorous examinations of the collaboration process. They touched on the description of the collaboration process through surveys and interviews with learners. The other cohort was articulately grounded on the LS research traditions, such as *collaboration for proximal outcome* and *collaboration as learning*, by designing AIs as *a Possibility engine* and a *Socratic Opponent* in a *Group-AI* collaboration context, and rigorously examining the collaboration process. Clusters 2 and 3 were more oriented toward *collaboration for proximal outcomes*, while cluster 3 was oriented toward *collaboration as learning*. Across all clusters, most studies focused on a specific analytical modality of collaboration, *Tactile*, because text-based communication has prevailed in the current AI development to support collaborative learning.

Our findings reveal that studies of AI in collaborative learning have been accelerated since 2020, but most strands have not yet examined collaboration processes in detail. They implemented AIs into personalized learning (*Human-AI* context) and primarily focused on distal learning outcomes, such as subject matter knowledge. One identified community is from CSCL and the LS, manifesting specific research traditions based on the epistemic stance of learning as knowledge-creation. However, the strand is not mature yet in designing AIs in the context of collaborative learning. Besides, most studies examined a specific type of analytical modality, *Tactile*, even though they implemented multimodal communication using AIs.

The future direction of studies will be directed toward the design of Group-AI collaboration supported by AIs' multimodal communication. Not only can AIs communicate with human learners in multimodal ways, but researchers also need to develop new theoretical orientations to design multimodal AI-supported CSCL environments and to analyze such dynamic communication.

Figure 1

Comparison Plots of Epistemic Networks Between Research Communities 3 and 6 (a), and 1 and 4 (b)



(a) (b)

Note. Square dots and dotted lines represent the centroids of research communities and confidence intervals, respectively.

References

Barron, B. (2003). When smart groups fail. *Journal of the Learning Sciences*, *12*(3), 307–359.

Enyedy, N. (2003). Knowledge construction and collective practice: At the intersection of learning, talk, and social configurations. *Journal of the Learning Sciences*, *12*(3), 361–408.

Enyedy, N., & Stevens, R. (2022). Analyzing collaboration. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (3rd ed., pp. 196–216). Cambridge University Press.

Esterhazy, R., Kaliisa, R., Sanchez, D., Langford, M., & Damșa, C. (2025). Advancing multimodal collaboration analytics: A scoping review. *Journal of Learning Analytics*, *12*(2), 105–124.

 © ISLS

Frederiksen, J. R., Sipusic, M., Sherin, M., & Wolfe, E. W. (1998). Video portfolio assessment: Creating a framework for viewing the functions of teaching. *Educational Assessment*, *5*(4), 225–297.

Hmelo-Silver, C. E., & Jeong, H. (2021). Benefits and challenges of interdisciplinarity in CSCL research: A view from the literature. *Frontiers in Psychology*, *11*, Article 579986.

Mende, S., Proske, A., & Narciss, S. (2021). Individual preparation for collaborative learning: Systematic review and synthesis. *Educational Psychologist*, *56*(1), 29–53.

Mikroyannidis, A., & Armakolas, S. (2025). Using generative AI to enhance personalized and self-regulated learning. In *EdMedia + Innovate Learning 2025* (pp. 249–258). Barcelona, Spain: Association for the Advancement of Computing in Education (AACE).

Nachtigall, V., Shaffer, D. W., & Rummel, N. (2022). Stirring a secret sauce: A literature review on the conditions and effects of authentic learning. *Educational Psychology Review*, *34*(3), 1479–1516.

Paavola, S., & Hakkarainen, K. (2005). The knowledge creation metaphor – An emergent epistemological approach to learning. *Science & Education*, *14*(6), 535–557.

Roschelle, J. (1992). Learning by collaborating: Convergent conceptual change. *Journal of the Learning Sciences*, *2*(3), 235–276.

Scardamalia, M., & Bereiter, C. (2022). Knowledge building and knowledge creation. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (3rd ed., pp. 385–405). Cambridge University Press.

Shaffer, D. W., Collier, W., & Ruis, A. R. (2016). A tutorial on epistemic network analysis: Analyzing the structure of connections in cognitive, social, and interaction data. *Journal of learning analytics*, *3*(3), 9–45.

Sharples, M. (2023). Towards social generative AI for education: Theory, practices and ethics. *Learning: Research and Practice*, *9*(2), 159–167.

Waltman, L., van Eck, N. J., & Noyons, E. C. M. (2010). A unified approach to mapping and clustering of bibliometric networks. *Journal of Informetrics*, *4*(4), 629–635.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Numbers 25K00836 and 25K21938.

 © ISLS

International Society of the Learning Sciences

Self-organizing Interdisciplinary Team in Generative AI-Supported Collaboration Context

Shotaro Naganuma, Kyushu University, naganuma.shotaro.062@m.kyushu-u.ac.jp
Tsubasa Minematsu, Kyushu Institute of Technology, minematsu@ai.kyutech.ac.jp
Jun Oshima, Shizuoka University, joshima@inf.shizuoka.ac.jp

Abstract: We report two-year design-based research (2024: N = 14, 2025: N = 19) on interdisciplinary research practice by doctoral students supported by generative AI (GenAI). The purpose was to explore necessary supports for interdisciplinary team formation by investigating how learners self-organized teams in this community. In both years, participants formed five teams to prepare interdisciplinary research proposals. Their video recordings data were analyzed. Findings revealed that AI-enabled exploratory collaboration experiences functioned as a collaboration history to build students' networks for the final team decision. We further identified three boundary-crossing patterns (e.g., Wide-Range Explorers). Through these diverse interactions, the community-level interdisciplinary distance converged toward a balance between novelty and mutual understandability. A more nuanced understanding of interdisciplinary team formation process in GenAI-supported setting, as well as the development of a visualization system displaying interdisciplinary distance, will benefit interdisciplinary team formation process.

Introduction

While today's complex societal problems demands higher education to provide interdisciplinary learning, several critical questions remain (Cai et al., 2025; Markauskaite et al., 2024). One of them is interdisciplinary team formation, requiring to seek collaborators outside of their academic networks (Daniel et al., 2022). Team formation has been commonly led either by instructors or by students. For example, mechanical and civil engineering students collaborated to solve a real industrial engineering problem (Bani-Hani et al., 2018) through random team assignment by the instructor or by students themselves. Another line of studies has provided technological support for team formation. It includes genetic algorithms or evolutionary algorithms, where students' knowledge, their learning styles, among others are used for grouping (e.g., Akhrif et al., 2021).

When applying in highly interdisciplinary contexts, two challenges can arise. First, predicting interdisciplinary collaboration fit in advance is a difficult task because participants have limited information about each other. Second, existing technological supports (e.g., genetic algorithms) can hardly work for the lack of any mutually comparable knowledge levels as a basis for grouping. Therefore, in our design-based research (DBR) project across two years, we attempted to provide learners with information that can be used for team formation through GenAI. Specifically, our students explored the communities through collaboration experiments, self-organized their own teams, and co-created interdisciplinary research proposals, with GenAI available to support them at every stage of the process. Therefore, by investigating how learners form their interdisciplinary teams in our GenAI-supported context, we aim to provide insights into designing and understanding more suitable team formation support in the future. The following research questions guided the study: *RQ1. How do the designed workshop for interdisciplinary research practice influence team formation? (macro) RQ2. During the team formation process, how do participants interact in the community? (micro)*

Theoretical framework

This study adopts the Complex Adaptive Systems (CAS) framework (Dooley, 1997; Holland, 1992) to conceptualize interdisciplinary team formation as an AI-augmented process of self-organization. Within CAS, multiple interacting agents generate emergent system-level structures through local interactions without centralized control. In our context, students function as agents who scan their classroom environment, engage in conversations, and construct interpretive and action-oriented schemas regarding potential collaborators. Through distributed and adaptive partner-selection processes, interdisciplinary teams emerge as socially negotiated outcomes of ongoing interaction. GenAI serves as an informational augmentation mechanism by projecting potential collaborative themes and expanding locally available decision-relevant information. Although team formation remains grounded in human interaction, AI-supported interaction may influence the trajectory of emergent team formations. This augmented CAS perspective shifts analytical attention from static individual attributes to relational dynamics, enabling examination of how conversational exchanges, adaptive decisions, and AI-mediated information jointly contribute to the emergence of interdisciplinary teams in classroom settings.

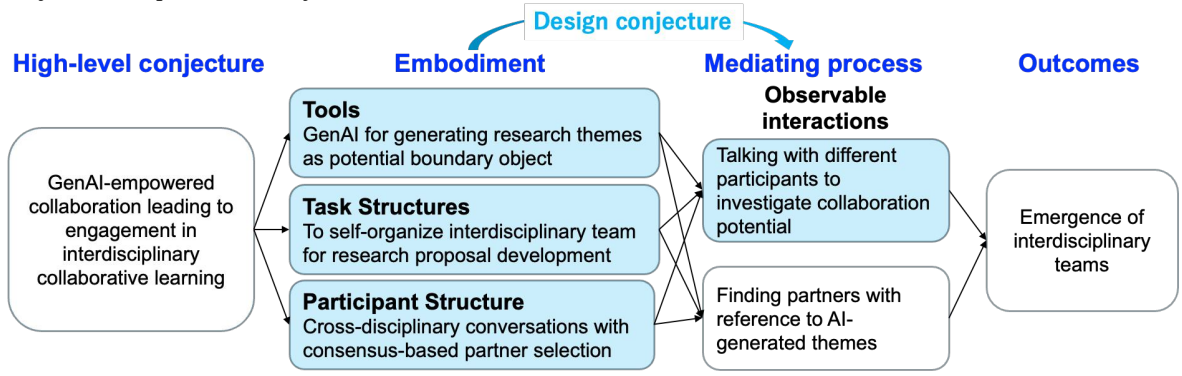
 © ISLS

Method

Our DBR (e.g., Brown, 1992) was conducted in 2024 and 2025 as three-day workshops. Figure 1 illustrates our conjecture map (Sandoval, 2014), showing how GenAI-empowered collaboration is expected to foster interdisciplinary engagement. The workshop design was refined based on our aforementioned pilot study (Naganuma et al., 2024; 2025; 2026). This DBR focuses on parts of the design conjecture concerning the relationship between embodiment and the expected mediating process toward team formation as outcomes.

Figure 1

Conjecture Map in This Study



Participants and contexts

Fourteen doctoral students at a national university in Japan participated in 2024, and 19 students in 2025, representing various departments from natural sciences to huminites. This study complied with ethical guidelines.

The workshop's goal was to develop high-quality interdisciplinary collaborative research proposals. It comprised eight 90-minutes sessions with four main activities. After the orientation (Sessions 1-2), students were assigned to teams and asked to judge the promisingness of AI-generated themes twice as exploratory collaboration experiments (Sessions 3-4). Second, they talked with others through four rounds (three 15 min-intervals and final round) to decide on the final team members (Session 5). After each interval, they rotated to talk with different students. Then, five teams were in 2024 and five teams were formed in 2025. Third, teams created research proposals (Sessions 6-7), and finally, each team gave a 10-minute presentation (Session 8).

Data collection and analysis

For RQ1, to investigate the implicit effect of incorporating collaboration experiment activities (Sessions 3 & 4), we used team composition record from the collaboration experiment activities and the final teams. We then visualized the team member transitions from two collaboration experiment activities to final team to investigate the implicit effect of collaboration experiment activities.

RQ2 focused on the team formation activity (Session 5), so video-recoding data of students' movement during the team formation process were collected. We then developed an *interdisciplinary distance* score adapted from the disciplinary affinity measure by Takahashi et al. (2023). The score was defined as 100 minus the affinity value between two disciplines. Since affinity is asymmetric, there are two distance values for each pair. For example, since the affinity from Neuroscience to Medicine is 58.2550 while the affinity from Medicine to Neuroscience is 3.2110, the corresponding interdisciplinary distances are 41.7450 and 96.7890, respectively. Calculations of the scores were based on 26 Scopus disciplines, excluding the General category. We analyzed the scores at two levels to capture both individual and collective aspects of interdisciplinary interactions.

For individual-level, we considered only the distances from their discipline to those of their partners, the values of which reflect the individual's interdisciplinary communication effort. For instance, if Member A talked with Members B and C, we averaged scores of A to B and A to C to quantify Member A's communication effort. After the calculation, to examine the individuals' interaction patterns toward team formation, we conducted hierarchical cluster analysis (Ward's method) on the interdisciplinary distance scores after eliminating missing values.

For team/community level, in each team, we calculated directed distances among all discipline pairs. For example, with three members (A, B, and C), six pairs exist (A to B, B to A, A to C, C to A, B to C, and C to B). Their average gave the team-level interdisciplinary distance scores. Then, we calculated the average of all teams' scores to represent the community-level interdisciplinary distance and also computed the standard deviation to capture how much the team-level indices varied within the community.

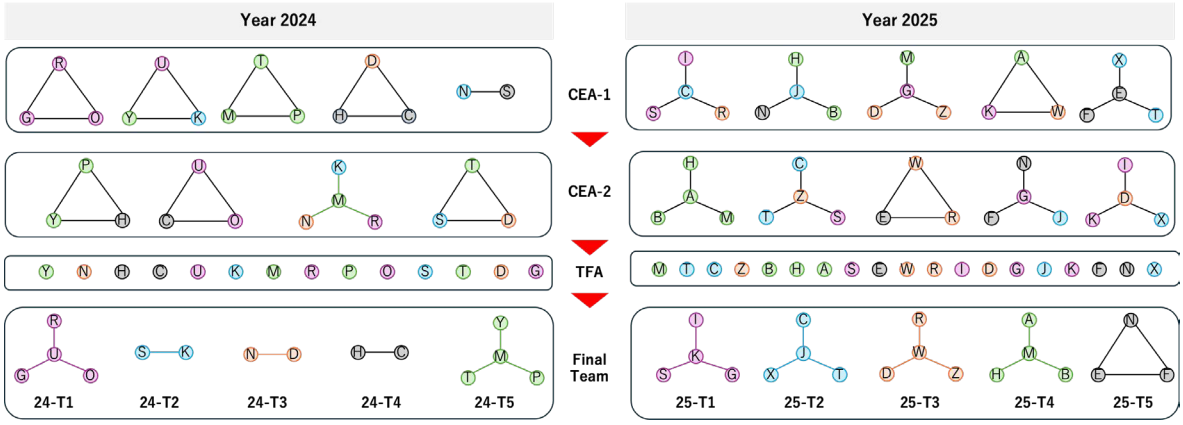
 © ISLS

Results and discussion

To address RQ1, figure 2 describes overall team transition patterns in 2024 and 2025. In 2024, three of the four members in two of the final team (Teams 24-T1 and 24-T5) had previously been grouped together in the same activity (CEA-1), and the two members who ultimately became a pair had also worked together in the CEA-1 (Team 24-T4). The 2025 practice showed the similar tendency. Prior studies have shown that individuals can share a common collaborator to form a new team (Liu et al., 2025) and innovative research teams are often composed of a mix of "members who have collaborated before" and "members working together for the first time" (Guimerà et al., 2005). This member mixture can provide both a sense of psychological safety and fresh perspectives, creating synergies. Our results clearly observed such team extensions. This suggests that collaboration experiment activities in Sessions 3 and 4 implicitly functioned as a collaboration history to help participants evaluate one another's collaboration potential toward team formations.

Figure 2

Overall Team Transition Pattern during the Workshop in 2024 and 2025



Note. CEA: Collaboration experiment activity; TFA: Team formation activity. Identical labels in different years indicate different individuals. Member G in 2024 was absent in CEA-2.

To address RQ2, using individual-level interdisciplinary distance scores, a cluster analysis identified three clusters in 2024: *Distant Anchors (N = 8)*, who consistently engaged in discussions with members from distant disciplines; *Boundary Wanderers (N = 4)*, who moved back and forth between moderately distant and widely distant disciplines; and *Wide-Range Explorers (N = 2)*, who traversed in the broad breadth of interdisciplinary distances in the community, ranging from very close to very distant disciplines. For different participants in 2025, we identified the three similar clusters. Table 1 shows that most final teams included participants from multiple clusters. Figure 3 illustrates three "*Wide-Range Explorers*" transitions during TFA.

Table 1

Participants' Distribution for the Three Boundary-crossing Patterns

Year	**2024**					**2025**				
Cluster	**24-T1**	**24-T2**	**24-T3**	**24-T4**	**24-T5**	**25-T1**	**25-T2**	**25-T3**	**25-T4**	**25-T5**
Distant Anchors	G, O, R, U	S	N		P, Y	I	C, T		A, H	N
Boundary Wanderers			D	C	M, T	G, S	J, X	D, W, Z	B, M	F
Wide-Range Explorers		K		H		K		R		E

Note. Identical labels in different years indicate different individuals.

Figure 4 illustrates the transitions of team- and community-level interdisciplinary distances during the TFA in 2024 and 2025. In the case of 2024, the figure highlights three salient characteristics. First, during the first 35 minutes, community-level interdisciplinary distances ranged from 70 to 90 with relatively low team-level variance, and this team-level variance tended to decrease over time. Second, in the next 20 minutes, team variance became large, which continued for approximately 10 minutes. Third, finally, the community-level

 © ISLS

interdisciplinary distance score converged to around 80, and team-level variance tended to decrease. Similar tendency was observed for the participants in 2025. Such visualization of community-level scores showed that community averages converged to 70–85 with reduced variance, suggesting that participants collectively sought balanced teams. Interpreted through the Optimal Cognitive Distance Hypothesis (Wuyts et al., 2005), this reflects a pursuit of both novelty and mutual understandability in creating original research proposals.

Figure 3

Three "Wide-Range Explorers" Transitions of Individual-level Interdisciplinary Distance

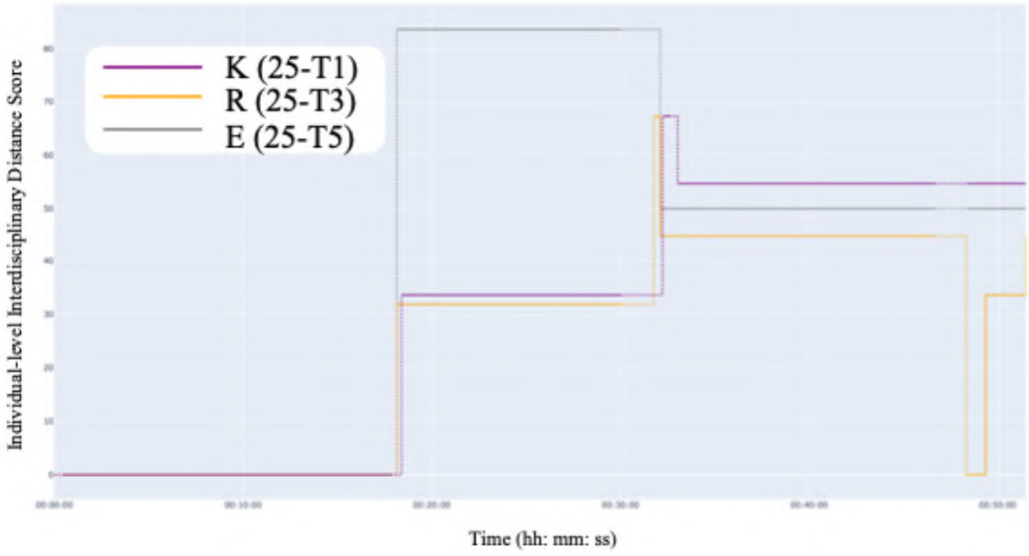
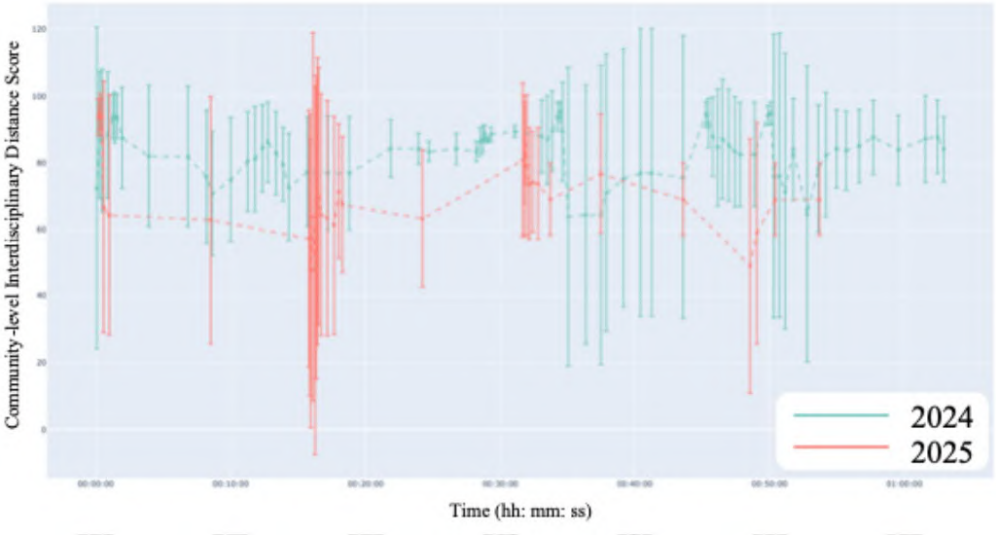


Figure 4

Transition of Team-level Interdisciplinary Distance Score and Team-level Variance



Conclusion

Using rich data from a two-year DBR project, we deepened our understanding of self-organizing process of interdisciplinary teams by doctoral students in the CAS augmented by GenAI, enriching our conjecture map (Figure 1). On the research side, future work will be benefited from analyzing conversation topics during team formation—e.g., personal topic, academic issue, social issues—including discussion on AI's ideas. On the development side, displaying real-time interdisciplinary distance within a community using technologies (e.g., Augmented Reality eyeglasses) might make interdisciplinary team formation more productive. While we used the data across two years, this study has been limited to an in-depth analysis of a few cases. Thus, we need to examine whether similar cases and patterns can be found in larger samples in other interdisciplinary learning programs.

 © ISLS

References

Akhrif, O., Benfaress, C., EL Jai, M., El Bouzekri El Idrissi, Y., & Hmina, N. (2021). Completeness based classification algorithm: A novel approach for educational semantic data completeness assessment. *Interactive Technology and Smart Education, 19*(1), 87–111. https://doi.org/10.1108/ITSE-01-2021-0017

Bani-Hani, E., Al Shalabi, A., Alkhatib, F., Eilaghi, A., & Sedaghat, A. (2018). Factors affecting the team formation and work in project based learning (pbl) for multidisciplinary engineering subjects. *Journal of Problem Based Learning in Higher Education, 6*(2), 136–143. https://eric.ed.gov/?id=EJ1200571

Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *Journal of the Learning Sciences, 2*(2), 141–178. https://doi.org/10.1207/s15327809jls0202_2

Cai, C., Zhu, G., & Ma, M. (2025). A systemic review of AI for interdisciplinary learning: Application contexts, roles, and influences. *Education and Information Technologies, 30*(7), 9641–9687. https://doi.org/10.1007/s10639-024-13193-x

Daniel, K. L., McConnell, M., Schuchardt, A., & Peffer, M. E. (2022). Challenges facing interdisciplinary researchers: Findings from a professional development workshop. *PLoS ONE, 17*(4), e0267234. https://doi.org/10.1371/journal.pone.0267234

Dooley, K. J. (1997). A Complex Adaptive Systems Model of Organization Change. Nonlinear Dynamics, Psychology, and Life Sciences, 1(1), 69–97. https://doi.org/10.1023/A:1022375910940

Guimerà, R., Uzzi, B., Spiro, J., & Amaral, L. A. N. (2005). Team assembly mechanisms determine collaboration network structure and team performance. *Science, 308*(5722), 697–702. https://doi.org/10.1126/science.1106340

Holland, J. H. (1992). Complex Adaptive Systems. *Daedalus, 121*(1), 17–30. https://www.jstor.org/stable/20025416

Liu, J., Tang, Z., Gong, X., Guo, X., & Zhang, L. (2025). Understanding the formation and dissolution of interdisciplinary teamwork networks: A comprehensive framework study of network structure, subject characteristics, and link attributes. *Journal of Informetrics, 19*(2), 101668. https://doi.org/10.1016/j.joi.2025.101668

Markauskaite, L., Schwarz, B., Damşa, C., & Muukkonen, H. (2024). Beyond disciplinary engagement: Researching the ecologies of interdisciplinary learning. *Journal of the Learning Sciences, 33*(2), 213–241. https://doi.org/10.1080/10508406.2024.2354151

Naganuma, S., Minematsu, T., Matsueda, K., & Oshima, J. (2024). Use of generative AI for boundary crossing in interdisciplinary collaborative research. In Clarke-Midura, J., Kollar, I., Gu, X., & D'Angelo, C. (Eds.), *Proceedings of the 17th International Conference on Computer-Supported Collaborative Learning - CSCL 2024* (pp. 67-74). International Society of the Learning Sciences.

Naganuma, S., Minematsu, T., Matsueda, K., & Oshima, J. (2025). Generative AI-supported collaboration in interdisciplinary research practice. In Oshima, J., Chen, B., Vogel, F., & Järvelä, J. (Eds.), *Proceedings of the 18th International Conference on Computer-Supported Collaborative Learning - CSCL 2025* (pp. 340-344). International Society of the Learning Sciences.

Naganuma, S., Minematsu, T., Matsueda, K., & Oshima, J. (2026). Doctoral students' interdisciplinary research proposal development supported by generative AI. *Education and Information Technologies,* Advance online publication. https://doi.org/10.1007/s10639-026-13952-y

Sandoval, W. (2014). Conjecture mapping: An approach to systematic educational design research. *Journal of the Learning Sciences, 23*(1), 18–36. https://doi.org/10.1080/10508406.2013.778204

Takahashi, R., Kaibe, K., Suzuki, K., Takahashi, S., Takeda, K., Hansen, M., & Yumoto, M. (2023). New concept of the affinity between research fields using academic journal data in Scopus. *Scientometrics, 128*(6), 3507–3534. https://doi.org/10.1007/s11192-023-04711-8

Wuyts, S., Colombo, M. G., Dutta, S., & Nooteboom, B. (2005). Empirical tests of optimal cognitive distance. *Journal of Economic Behavior & Organization, 58*(2), 277–302. https://doi.org/10.1016/j.jebo.2004.03.019

Acknowledgement

This work was supported by JSPS KAKENHI Grant Number 25K00836.

International Society of the Learning Sciences

Exploring Shared Agency in an Intelligent Immersive Learning Environment for Collaborative Writing

Yun Wen, Nanyang Technological University, Singapore, yun.wen@nie.edu.sg
Zongran Wang, East China Normal University, China, wangzongran0215@gmail.com
Lang Xu, Nanyang Technological University, Singapore, nie25.xl@e.ntu.edu.sg

Abstract: An increasing number of studies on AI-student collaboration highlight the importance of learner agency in the Generative AI (GenAI)-powered learning environment. However, limited research has examined how student-AI shared agency can be assessed, and it remains unclear whether individual interaction with AI agents or peer collaboration with AI agents is more effective in developing such agency. This study explores second language (L2) students' shared agency within an intelligent immersive learning environment for writing. Using WriteCube, a self-developed AR- and AI-enabled writing system, 75 Primary 4 students from two classes in Singapore completed picture-based composition tasks under individual-AI and pair-AI collaboration conditions. Findings revealed no significant performance difference between groups, suggesting a trade-off between interactional richness and task efficiency. Regression results indicated that interaction quality, rather than frequency or language proficiency, positively predicted shared agency, particularly in ensuring equity. The study offers implications for assessing and enhancing learner agency in GenAI-supported L2 classrooms.

Introduction

Agency has become a central concern in discussion on Generative AI (GenAI) integration in teaching and learning (Darvishi et al., 2024). The notion of agency has been typically defined as the learner's capacity to act intentionally and exert control over their learning (Bandura, 2002). While from the sociocultural perspective, agency is not merely a personal trait but also an inherently collective phenomenon, shaped by social, cultural, and contexts (Eteläpelto et al., 2013), and fundamentally emerging as the result of inter-subjective processes. As the use of GenAI increasingly destabilizes boundaries between humans and technology, it becomes critical to examine learner agency by considering how they navigate, critique, and shape learning experiences with GenAI (Code, 2025). This study seeks to examine shared agency between students and AI agents, building on the concept of collective human agency in studies of computer-supported collaborative learning (Scardamalia & Bereiter, 2014; Zhang et al., 2021), while also incorporating considerations of ethical engagement and critical awareness (Ralston, 2023). Student-AI shared agency, in this study, is conceptualized as the emergent and dynamic capacity of students to co-act with AI agents. There remains a lack of research on how shared agency can be assessed (Code, 2025). Moreover, it is still unclear whether individual interaction with AI agents or peer collaboration with AI agents is more effective in developing Student-AI shared agency.

To address this gap, the present study was proposed and conducted based on WriteCube, a self-designed and developed writing system for Chinese as a second language (L2) learning. The system brings students into immersive augmented reality (AR) scenarios, empowered by AI agents. AR and AI together offer complementary strengths for designing L2 writing environments (Wen et al., 2025). While AI support often remains decontextualized, lacking the authentic and multimodal cues needed to ground language use in real-world contexts, AR technologies provide authentic, multisensory, and context-rich environments that situate writing tasks in meaningful scenarios, enhance motivation, and reduce language anxiety (Yang et al., 2023). The system was implemented in two Primary 4 Chinese language classes from a school in Singapore. This study aimed to deepen understanding and empirically examine agency in the GenAI-powered learning environments. The findings sought to provide design and implementation suggestions for using such learning systems in classrooms.

Theoretical foundations

Code developed the Agency for Learning Framework to measure agency in learning context by considering both individual and environmental influences on learners' ability to act (Code, 2020). As the agency of technology is acknowledged, technology does not only play a role in mediating learner agency. Code further proposed the updated framework about the Postdigital Learner Agency, "equipping leaners to navigate, critique, and shape the systems mediating their educational experiences" (Code, 2025, p.338). In this perspective, agency extends beyond individual action to include collective and relational dimensions. Collective agency emphasizes collaboration, co-regulation, and shared goals, fostering mutual understanding and problem-solving (Lipponen & Kumpulainen

 © ISLS

2011), while relational agency emphasizes the dynamic and contextual nature of agency shaped through interactions with both human and non-human actors (Siles et al., 2024). Building on this perspective, the present study proposes student-AI shared agency, highlighting the dynamic and contextual nature of agency emerging through interactions between learners and AI systems.

The integration of GenAI as a partner broadens the scope of collaborative L2 writing. Existing studies have examined how students individually use GenAI in collaborative writing contexts (e.g., Kim et al., 2025), as well as how human peers work with GenAI to mitigate limitations in language proficiency that may hinder effective interaction with AI agents (e.g., Michel et al., 2025). Human peers bring contextual sensitivity, affective support, and stronger benefits for long-term retention (Yang et al., 2023). Despite growing research on AI-assisted writing, little is known about how factors such as students' language proficiency, engagement level, and collaboration mode (individual-AI vs. human peer collaboration) shape student-AI shared agency in L2 writing. Addressing this gap is essential for understanding how human and AI collaborators jointly contribute to the development of shared agency in AI-supported writing environments. The following research questions guide this pilot study.

RQ1: Do individual and pair writing modes differentially impact students' writing performance and student-AI shared agency development?

RQ2: What factors are associated with student-AI shared agency development in the intelligent immersive learning environment?

Research design

Participants and setting

The study involved 75 Primary 4 students from two intact classes in a Singapore primary school, taught by the same Chinese language teacher (15 years' experience). The two classes were selected because they followed the same curriculum and were taught by the same teacher, helping hold instructional conditions constant while implementing different collaboration modes. Class A worked in the individual-AI condition (N = 37). Class B worked in the pair-AI condition (N = 38), with pairs randomly assigned within the class.

Procedure

Students completed a 45-minute technical familiarization session followed by a 60-minute WriteCube activity. WriteCube renders task images as AR scenes and supports composing through two interaction phases: (1) a touchpoint phase for fine-grained description via short dialogues and role-play, and (2) a chatroom phase for developing coherent paragraphs when additional support is needed. An AI agent prompts idea elaboration, recommends expressions, and provides feedback. In a final stage, the AI generates a draft composition from students' inputs and students iteratively revise and complete a self-reflection checklist.

Measures

(a) Writing performance: pre- and post-writing compositions were scored on the school's 20-point rubric by two trained raters; final scores were averaged. We tested condition differences via one-way ANCOVA with pre-writing as covariate. (b) Shared agency: a post-survey adapted from a collective agency scale (Zhang et al., 2022) and extended with items reflecting reflective/ethical engagement with AI (Ralston, 2023). Exploratory factor analysis (Varimax) yielded three indices after removing two cross-loading items: ***autonomy perception*** (2 items, α = .63, e.g., I try to think of my own/my group sentences and ideas before reading what the AI writes.), ***collective efficacy*** (7 items, α = .87, e.g., I believe I/my group can work together with AI to solve writing problems.), ***ensuring equity*** (5 items, α = .86, e.g., Before starting the writing, I think about what I/my group want to write by myself/ourselves and where I/we want the AI to help).Condition effects were examined via MANOVA.

Process indicators and analysis:

Dialogue logs captured full student-AI exchanges. We computed (i) number of turns, (ii) average turn length, and (iii) number of 'good words' (teacher-endorsed descriptive keywords recorded in system logs) as a proxy of interaction quality. Multiple regressions examined associations between shared-agency indices and pre-writing scores plus process indicators.

Findings

No significant differences in either writing performance or shared agency

ANCOVA showed no significant difference in post-writing scores between the individual-AI and pair-AI conditions, $F(1, 58) = 3.65$, $p = .061$, partial $\eta^2 = .059$, although the individual-AI group showed a slightly higher adjusted mean (12.13 vs. 10.46). Process logs indicated that students in the individual-AI condition produced more dialogue turns (38.45 vs. 26.71), whereas turns in the pair-AI condition were slightly longer on average (9.26 vs. 8.48 words); the number of "good words/keywords" was comparable across groups (Table 1).

Box's test of equality of covariance matrices was examined ($p > 0.05$), indicating that the equality of covariance matrices was met. MANOVA further suggested no significant overall effect of condition on the three shared-agency dimensions, Wilks' $\Lambda = .958$, $F(3, 54) = 0.80$, $p = .502$, partial $\eta^2 = .042$. Follow-up univariate tests were also nonsignificant for autonomy perception, collective efficacy, and ensuring equality (Table 2), with ensuring equality descriptively higher in the pair-AI group.

Table 1
Descriptive Analysis of Student-AI Dialogues

Explanatory variable	**Mean (SD)**	
	Individual -AI	Pair-AI
Number of turns	38.45 (9.21)	26.71 (11.49)
Length of each turn	8.48 (4.73)	9.26 (6.67)
Number of good words	3.30 (1.78)	3.06 (1.86)

Table 2
MANOVA Summary of Student-AI Shared Agency

Dependent Variable	***F*(1, 56)**	***p***	***Partial η²***	**Mean (SD)**	
				Individual -AI	**Pair -AI**
Autonomy perception	0.09	.769	.002	3.60 (.19)	3.67 (.17)
Collective efficacy	0.50	.482	.009	3.87 (.18)	3.70 (.16)
Ensuring equity	0.42	.521	.007	3.45 (.20)	3.63 (.18)

Factors associated with student-AI shared agency development

Variables showed low intercorrelations and were retained for regression analyses. Separate regressions indicated that only the model predicting ensuring equality reached significance. Specifically, the number of good words/keywords positively predicted ensuring equality ($b = .191$, $p < .05$), whereas the number of dialogue turns negatively predicted it ($b = -.022$, $p < .05$), explaining 18% of variance ($R^2 = .180$). Models for autonomy perception and collective efficacy were not significant.

Qualitative inspection of dialogue logs suggests that many high-frequency exchanges reflected prompt-following behavior, where students adopted AI suggestions with limited elaboration. For example, students often directly reused AI-provided expressions such as "眉头紧皱" in expanded descriptions (e.g., "丽丽低着头，眼睛看着桌子，眉头紧皱"), or inserted idiomatic phrases like "心里像吃了蜜糖一样甜" with minimal modification. In contrast, episodes associated with higher perceived equality involved meaning clarification and integration, such as a student asking "尽管是什么意思？" and then applying it appropriately in the next sentence. Some students also treated the AI as a linguistic resource, requesting support (e.g., "你能给我一些好词好句吗？") before integrating the expressions into their narratives. Notably, explicit critique of AI suggestions was rarely observed.

Discussion

The absence of significant performance differences between the individual-AI and pair-AI groups suggest a potential trade-off between interactional richness and task efficiency. Students working individually interacted with the AI more frequently and maintained continuous task engagement, which may explain their slightly higher writing scores. In contrast, pair-AI collaboration involved fewer but longer dialogue turns, reflecting additional peer negotiation. This pattern aligns with research suggesting that early-stage L2 writers often move from imitative uptake toward deeper linguistic integration as they internalize scaffolding (Guirao et al., 2015). While peer collaboration may promote reflection and evaluation, such benefits may not immediately appear in short-term performance measures.

Regression results further indicated that interaction quality rather than interaction frequency predicted shared agency, particularly in the dimension of ensuring equality. Students who incorporated more meaningful expressions into their dialogues reported stronger perceptions of balanced collaboration with AI, whereas more frequent but surface-level exchanges were associated with lower perceived equality. Interestingly, language

proficiency did not significantly predict shared agency. This suggests that, compared with traditional peer collaborative writing where proficiency often shapes participation (Susanti et al., 2020), AI-supported environments may reduce participation disparities and allow lower-proficiency learners to engage more actively.

Overall, these findings highlight that students' perception of shared agency depends less on how often they interact with AI and more on how meaningfully they integrate AI suggestions into their own ideas. Future research should further explore instructional strategies that encourage deeper interaction with AI, including prompting students to question, reinterpret, or critique AI-generated suggestions.

Conclusion

This exploratory study examined student-AI shared agency in an intelligent immersive L2 writing environment. The findings suggest that individual-AI and pair-AI collaboration produced comparable writing outcomes and similar levels of shared agency, while interaction quality emerged as a key factor associated with students' perceptions of equal collaboration with AI. Several limitations should be acknowledged. The relatively small sample size and short intervention period may have limited the statistical power to detect subtle differences between conditions. Future studies with larger samples and longitudinal designs are needed to examine longer-term effects on language development and agency formation. In addition, while qualitative observations helped interpret quantitative patterns, a more systematic analysis of student-AI interaction processes could further illuminate how shared agency develops over time. Despite these limitations, the study provides an initial empirical perspective on assessing shared agency in AI-supported learning environments and highlights the importance of designing learning activities that foster reflective and meaningful human-AI interaction.

Reference

Bandura, A. (2002). Selective moral disengagement in the exercise of moral agency. *Journal of moral education*, *31*(2), 101-119.

Code, J. (2020). Agency for learning: Intention, motivation, self-efficacy and self-regulation. *Frontiers in Education*, *5*, 19. https:// doi. org/ 10. 3389/ feduc. 2020. 00019.

Code, J. (2025). The entangled learner: Critical agency for the postdigital era, *Postdigital Science and Education, 7*, 336-358.

Darvishi, A., Khosravi, H., Sadiq, S., Gašević, D., & Siemens, G. (2024). Impact of AI assistance on student agency. *Computers & Education, 210*, 104967.

Eteläpelto, A., Vähäsantanen, K., Hökkä, P., & Paloniemi, S. (2013). What is agency? Conceptualizing professional agency at work. *Educational Research Review, 10*, 45–65.

Guirao, J. C., de Larios, J. R., & Coyle, Y. (2015). The use of models as a written feedback technique with young EFL learners. *System*, *52*, 63-77.

Kim, J., Yu, S., Lee, S. S., & Detrick, R. (2025). Students' prompt patterns and its effects in AI-assisted academic writing: Focusing on students' level of AI literacy. *Journal of Research on Technology in Education*, 1–18.

Lipponen, L., & Kumpulainen, K. (2011). Acting as accountable authors: Creating interactional spaces for agency work in teacher education. *Teaching and teacher education*, *27*(5), 812-819.

Michel, M., Bazhutkina, I., Abel, N., & Strobl, C. (2025). Collaborative writing based on generative AI models: Revision and deliberation processes in German as a foreign language. *Journal of Second Language Writing, 67*, 101185.

Ralston, S. (2023). Postdigital ethics. In P. Jandrić (Ed.), Encyclopedia of postdigital science and education. Springer. https:// doi. org/ 10. 1007/ 978-3- 031- 35469-4_ 49-1.

Scardamalia, M., & Bereiter, C. (2014). Knowledge building and knowledge creation: Theory, pedagogy, and technology. In K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (2nd ed., pp. 397–417). Cambridge University Press. https://doi.org/10.1017/CBO9781139519526.025

Siles, I., Gómez-Cruz, E., & Ricaurte, P. (2024). Fluid agency in relation to algorithms: Tensions, mediations, and transversalities. *Convergence*, *30*(3), 1025-1040.

Susanti, A., Widiati, U., & Cahyono, B. Y. (2020). The effect of proficiency pairings on EFL students' writing ability in genre-based approach context. *International Journal of Evaluation and Research in Education*, *9*(1), 245-251.

Wen, Y., Wang, Z., & Guo, X. (2025). Trends and applications of AI in immersive learning environments: a systematic review of empirical research. *Interactive Learning Environments*, 1–19. https://doi.org/10.1080/10494820.2025.2524029

Yang, Y. F., Goh, A. P., Hong, Y. C., & Chen, N. S. (2023). Primary school students' foreign language anxiety in collaborative and individual digital game-based learning. *Computer Assisted Language Learning*, *36*(8), 1587-1607.

Zhang, S., Wen, Y., & Liu, Q. (2022). Exploring student teachers' social knowledge construction behaviors and collective agency in an online collaborative learning environment. *Interactive learning environments*, *30*(3), 539-551.

Acknowledgments

This study is funded by Ministry of Education (MOE) Social Science and Humanities Research (SSHR) Fellowship for the project MOE SSHRF 8/22 WY [MOE2021-SSHR-009].

 © ISLS

International Society of the Learning Sciences

Student Perceptions of Generative AI Chatbots in Computing Education: A Technology Acceptance Lens

Ean Teng Khor, Elizabeth Koh, Leta Chan
eanteng.khor@nie.edu.sg, elizabeth.koh@nie.edu.sg, leta.chan@nie.edu.sg
National Institute of Education, Nanyang Technological University, Singapore

Abstract: As generative artificial intelligence (GenAI) becomes increasingly embedded in educational settings, understanding students' acceptance of such tools is critical. Grounded in the Technology Acceptance Model (TAM), this study investigates secondary students' perceptions of *MyBotBuddy*, a customized GenAI programming chatbot designed to scaffold learning. 49 computing students aged 15 to 16 completed a programming task before and after interacting with the chatbot and responded to an open-ended survey. Although quantitative results showed no statistically significant improvement in programming scores following the intervention, qualitative findings revealed generally positive perceptions. Students valued clarity, structured explanations, scaffolding support, and convenience. Perceived usefulness was shaped primarily by reliability, academic benefits, and system capability, while perceived ease of use was influenced by response clarity and the effort required to interact with the system. The findings highlight output quality, adaptability, response speed, and accuracy as central to educational chatbot acceptance. This study translates TAM constructs into practical design implications for GenAI tools in K–12 computing education.

Introduction

GenAI tools have rapidly gained prominence in educational contexts. These systems offer the potential to personalize instruction, provide immediate feedback, and support learners beyond classroom hours (Wu et al., 2023). In programming education, GenAI can generate code, diagnose errors, and propose alternative approaches to problem-solving (Zhang et al., 2024). From a computer-supported collaborative learning perspective, dialogue between learners and AI may also facilitate co-construction of knowledge (Cress & Kimmerle, 2023). Despite these affordances, the successful integration of GenAI into classrooms depends not only on technical capability but also on student acceptance. The TAM developed by Davis (1989) posits that perceived usefulness and perceived ease of use are key determinants of user adoption. Extensions of TAM further identify output quality, relevance, system characteristics, and experience-related factors as influencing these perceptions (Venkatesh & Davis, 2000). Existing research frequently investigates students' perceptions of general-purpose chatbots such as ChatGPT, yet fewer studies examine customized educational chatbots designed for specific curricular contexts. Moreover, limited work explicitly connects concrete chatbot design features to TAM constructs in K–12 settings. This study addresses this gap by examining students' perceptions of *MyBotBuddy*, a GenAI chatbot developed to scaffold secondary-level programming learning (Khor et al., 2026). The study asks: To what extent is *MyBotBuddy* perceived as useful and easy to use, and what factors shape these perceptions?

Literature review

The TAM conceptualizes perceived usefulness as the degree to which a user believes a system enhances performance, and perceived ease of use as the degree to which it is free of effort (Davis, 1989). TAM2 and TAM3 elaborate that output quality, job relevance, social influence, and system characteristics significantly shape these perceptions (Venkatesh & Davis, 2000). In the context of GenAI, research consistently reports that students perceive AI chatbots as useful for improving efficiency, clarifying concepts, and simplifying coding tasks (Baidoo-Anu et al., 2024). Ease of use is often attributed to conversational interfaces, accessibility, and well-structured explanations (Vo & Nguyen, 2024). However, concerns regarding hallucinations, misinformation, and overreliance remain prevalent (Zastudil et al., 2023). Reliability has been shown to significantly affect trust and perceived usefulness (Weisz et al., 2022). While much of the literature focuses on higher education and general chatbots, there is comparatively limited investigation into customized K–12 educational tools and how their specific design features influence TAM constructs. Designing GenAI learning systems requires attention not only to functionality but also to scaffolding, cognitive load, and adaptability. Clear and structured explanations reduce extraneous cognitive load and support learning (Sweller, 1988). Furthermore, GenAI systems should encourage active engagement rather than passive answer retrieval (Cress & Kimmerle, 2023). Despite these insights, empirical work linking such design features to perceived usefulness and ease of use remains underdeveloped. This study therefore integrates TAM with concrete system features to better understand student acceptance.

 © ISLS

Research design and methods

A design-based research approach guided the development and evaluation of *MyBotBuddy* (Khor et al., 2025). The chatbot was informed by Vygotsky (1978)'s Zone of Proximal Development and implemented a scaffolding model that provided tiered support, beginning with conceptual explanation, followed by examples, and culminating in diagnostic feedback (Khor & Chan, 2025). The system incorporated retrieval-augmented generation aligned with curricular materials, as well as a secure code execution sandbox to analyze student submissions and return structured feedback. Responses were segmented into manageable sections to reduce cognitive load, and safeguards were implemented to ensure age-appropriate and curriculum-aligned outputs. 49 secondary school computing students aged 15 to 16 participated in the study. Students completed a 30-minute Python programming task involving ISBN validation both before and after a one-hour intervention using *MyBotBuddy* (Khor et al., 2024). During the intervention, students were encouraged to use the chatbot to refine their code, troubleshoot errors, or explore alternative solutions. Quantitative data were collected through pre- and post-test scores graded with a standardized rubric. Three cases were excluded due to technical issues, resulting in 46 valid paired datasets. A paired-samples t-test was conducted to assess differences in performance. Qualitative data was collected through an open-ended survey administered after the intervention. Thematic analysis following Naeem et al. (2023) was employed, with codes deductively organized under TAM constructs of perceived usefulness and perceived ease of use.

Findings

The paired-samples t-test showed no statistically significant improvement in students' programming performance from pre-test ($M = 6.26$, $SD = 3.42$) to post-test ($M = 6.87$, $SD = 3.10$), $t(45) = -1.176$, $p = .123$, 95% CI [−1.65, .43], suggesting that short-term interaction with *MyBotBuddy* did not yield measurable gains in performance. This indicates the need for further refinement to strengthen its objective instructional impact. In contrast, qualitative findings reflected generally positive perceptions. Many students highlighted the chatbot's support in understanding programming concepts, and several spontaneously compared it with ChatGPT. While some perceived both systems as comparable, others preferred *MyBotBuddy* for its clearer explanations. Thematic analysis, guided by TAM3, identified four key themes. Reliability, academic benefits, and system capabilities shaped perceived usefulness, while the effort required to use the chatbot influenced perceived ease of use. These themes informed directions for further system improvement.

Perceived ease of use

Students generally agreed that *MyBotBuddy* was easy to use, emphasizing its convenience, clarity, and well-structured responses. *MyBotBuddy*'s responses were described as "comprehensive and simple" (S28) and "concise yet easy to understand" (S9). They noted that the step-by-step breakdowns helped them grasp algorithms without having to read extensive code. The use of uncomplicated language, highlighted keywords, and clearly segmented sections reduced cognitive effort and made key ideas easy to identify. However, preferences varied. While some students valued concise responses, others appreciated more detailed explanations. This divergence suggests that adaptability is central to perceived ease of use. Students may expect a GenAI learning tool to provide adjustable levels of explanation to suit individual needs. Although overall perceptions were positive, the lack of consensus on specific improvements indicates that perceived ease of use may depend less on a single optimal design and more on the system's capacity to accommodate diverse user preferences.

Perceived usefulness

Students' perceptions of *MyBotBuddy*'s usefulness centered on three main aspects: reliability, academic benefits, and system capability. Reliability was particularly salient. Although most students found the chatbot generally accurate, for instance, "much better and well-structured compared to other AI models" (S30), occasional logic errors in generated code raised concerns. While some noted that the system could correct itself after further interaction, accuracy remained a key expectation. Students viewed reliable, high-quality responses as essential for supporting learning, and inaccuracies were seen as potentially confusing or detrimental. These findings suggest that strengthening output accuracy is critical to enhancing perceived usefulness, particularly for an education-oriented chatbot intended to outperform general-purpose tools in domain reliability.

Students also reported substantial academic benefits. *MyBotBuddy* was perceived as helpful in refining code, identifying and explaining errors, guiding program design, and suggesting alternative solutions. Students appreciated how *MyBotBuddy* would provide "clear step by step explanations" (S42), "gives the explanation of the code and constraints and the approach and what the variables do" (S30), and even provide "useful tips and

 © ISLS

revisions to help me realize what is wrong and how to improve" (S49). This suggests that the scaffolding approach of the chatbot were valued for learning, particularly by beginners, and supported conceptual understanding and encouraged creative problem-solving. These features contributed directly to their perceived usefulness as a programming learning aid.

However, perceived usefulness was constrained by limitations in scope and system performance. Because *MyBotBuddy* was restricted to programming-related queries, some students viewed it as overly specialized and less versatile than broader AI systems. They expressed a preference for tools applicable across subjects, suggesting that cross-disciplinary utility influences perceived relevance. Students also expected a system that could handle complex inputs and adapt to varied prompts. Enhancing technical robustness and expanding functional scope may therefore strengthen perceptions of usefulness and support deeper learning engagement.

Discussion

This study aimed to identify the determinants of students' perceptions of *MyBotBuddy* using the technology acceptance lens to provide insights into *MyBotBuddy*'s current strengths and limitations and suggest improvements. One of *MyBotBuddy*'s greatest strengths as mentioned by the students is its **clear, detailed, and step-by-step guidance** when explaining programming concepts and code that facilitates students' understanding of programming concepts. This feature forms the basis of its perceived utility to students – the bot's ability to effectively scaffold students' programming knowledge. This aligns with previous literature that highlights how GenAI chatbots' ability to explain learning materials clearly and thoroughly influence students' opinions on these tools (Shoufan, 2023) as it simplifies their studying process by providing relevant, easily accessible, and understandable information (Kayali et al., 2023).

Output quality, or students' belief that *MyBotBuddy* performs its role as a tutor well, is theorized in TAM as a primary driver of perceived usefulness (Venkatesh & Davis, 2000). When *MyBotBuddy* delivers clear, specific, and accurate explanations, students infer that it effectively fulfils the role of a programming tutor and, in turn, find it more useful. As earlier shown, our qualitative comments foreground *explanation quality* and align with this mechanism. This suggests that refining the chatbot's scaffolding and accuracy can raise perceived output quality and thereby strengthen perceived usefulness (Venkatesh & Davis, 2000).

Another strength of *MyBotBuddy* is its **concise, uncomplicated and well-structured answers**. In line with Sweller's (1988) cognitive load theory, these features could remove obstacles, such as complex jargon or overwhelmingly long text, that may interfere with students' learning by competing for limited cognitive resources in understanding them. This reduces the cognitive load students face when interacting with *MyBotBuddy* and frees up cognitive resources for internalizing the content learned and problem-solving. Qualities of students' experience while using a learning tool have been proposed to influence their perceptions on how easy it is to use in the TAM3 model (Venkatesh & Davis, 2000). The model theorizes that user-friendly system characteristics encourage students' feelings of control while using it, allowing them to feel more competent and perceive it as being easier to use. Thus, it suggests that the simplicity of *MyBotBuddy*'s answers supports students' confidence to use *MyBotBuddy*.

The above strengths both refer to specific characteristics of *MyBotBuddy*'s generated responses, suggesting that the extent of its perceived usefulness and ease of use heavily relies on the quality of its responses. The implication is that the chatbot's perceived usefulness and ease of use can thus be maximized by ensuring that it generates high quality responses which are clear, thorough, detailed, simple, and neatly structured. To further improve *MyBotBuddy*'s utility and ease of use, it is imperative to address its key weaknesses, namely its slow response speeds, issues with reliability, and narrow field of expertise. An interesting finding was that students preferred more general chatbots when studying and suggested that it may be more useful in aiding their learning than a specialized chatbot. Instead of using a general chatbot, *MyBotBuddy* intentionally focused on teaching programming to lower risks of AI hallucinations or propagating bias as it has the added advantage of being able to control the data that it was trained on to ensure its reliability (Yeung et al., 2023). Educational chatbots designed to teach a specific field of study, like the Anatbuddy chatbot by Arun et al. (2024), were more factually accurate than their general counterparts like ChatGPT. Students' view of the subject-specific nature of *MyBotBuddy* as a limitation was thus surprising. This preference may reflect students' preference for convenience: instead of switching between different learning resources according to each subject, students may prefer seeking help from a central source. The literature supports this, suggesting that students were driven to use ChatGPT while studying due to how it saves time and effort in studying (Jishnu et al., 2023). As theorized in the TAM3, learning tools which are seen as being less effortful to use may be found more useful, thus expanding the scope of *MyBotBuddy*'s knowledge may improve its perceived usefulness to students (Venkatesh & Davis, 2000). Alternatively, this preference may be attributed to students' perceptions of *MyBotBuddy*'s relevance. Students may judge a learning

 © ISLS

tool's relevance according to its overall applicability to their studies, rather than to a specific subject. A learning tool which has academic benefits extending beyond a single subject but to one's overall studies may be considered more relevant to their learning and hence seen as more useful (Venkatesh & Davis, 2000). Based on this preference for an integrated platform, it would be beneficial to expand the scope of the tool's knowledge or include additional resources related to students' studies to maximize its perceived usefulness. **The extent of *MyBotBuddy*'s perceived usefulness may thus be tied to its ability to effectively meet students' specific needs and learning goals.** This highlights the ongoing challenge for educational GenAI tools to stay relevant and effective by continuously updating and integrating the latest advancements in the field (Chukwuere & Handoko, 2024).

MyBotBuddy's weakness in consistently generating accurate answers is closely tied to students' perceived usefulness of it, as often a lack of consistent accuracy in AI-generated answers undermines their trust in the tool's value to effectively assist them (Weisz et al., 2022). Inaccurate information could breed misconceptions, which would hinder students' success in programming and cause students to grow wary of *MyBotBuddy*'s usefulness as a learning tool (Shoufan, 2023). Students' **concern over the accuracy of *MyBotBuddy*'s responses** may reflect the influence of perceived output quality on the perceived usefulness of a tool, as theorized in the TAM3 (Venkatesh & Davis, 2000). Students who received inaccurate generated answers may negatively view how well *MyBotBuddy* performs its role as a tutor, leading them to find it less relevant to their learning and thus less useful to them as a learning tool. The perceived usefulness of an educational chatbot can thus be enhanced by targeting output quality, ensuring its responses are both reliable and accurate.

The non-significant quantitative findings of the test scores in this intervention as mentioned could be due to the short nature of the task; also as with many educational interventions, students need time to get used to learning in a new approach and short interventions could indicate tool-fit more than learner growth (Arun et al., 2024). Furthermore, human-AI inquiry seeks knowledge building through argumentative dialogue, which is inherently iterative and time-consuming relative to didactic methods (Cress & Kimmerle, 2023). Still the qualitative findings suggest that several students did learn through their scaffolding provided in their interaction with *MyBotBuddy*.

Conclusion

The present study examined students' perceptions of *MyBotBuddy*, a GenAI programming chatbot, through a technology acceptance lens to inform system improvement. Qualitative findings indicate that perceived usefulness was shaped primarily by response quality, demonstrated expertise, and accuracy. Perceived ease of use depended largely on how concise, simple, and well-structured the responses were. Students' suggestions for improvement focused mainly on enhancing output quality and interactivity, highlighting these as central levers for strengthening acceptance. Improving feedback clarity, reliability, and responsiveness may enable students to engage in deeper, more meaningful dialogue with the chatbot, moving beyond passive answer retrieval toward more collaborative human–AI problem-solving. However, the study relied on an open-ended survey and a relatively short interaction period, which may have limited students' depth of engagement and shaped their perceptions. Longer-term implementation could provide richer insights into how perceptions evolve over time. Another limitation of the study is the lack of an alternative condition to compare the effects of the intervention. Nonetheless, the findings translate theoretical constructs of technology acceptance into practical design recommendations to support more effective and engaging educational GenAI tools.

References

Arun, G., Perumal, V., Urias, F. P. J. B., Ler, Y. E., Tan, B. W. T., Vallabhajosyula, R., Tan, E., Ng, O., Ng, K. B., & Mogali, S. R. (2024). ChatGPT versus a customized AI chatbot (Anatbuddy) for anatomy education: A comparative pilot study. *Anatomical Sciences Education, 17*(7), 1396-1405. https://doi.org/10.1002/ase.2502

Baidoo-Anu, D., Asamoah, D., Amoako, I., & Mahama, I. (2024). Exploring student perspectives on generative artificial intelligence on higher education learning. *Discover Education, 3*, Article 98. https://doi.org/10.1007/s44217-024-00173-z

Chukwuere, J. E., & Handoko, B. L (2024). The future of generative AI chatbots in higher education. *Journal of Emerging Technologies*, *4*(1), 36-44. https://doi.org/10.57040/y62k8d79

Cress, U., & Kimmerle, J. (2023). Co-constructing knowledge with generative AI tools: Reflections from a CSCL perspective. *International Journal of Computer-Supported Collaborative Learning, 18,* 607-614. https://doi.org/10.1007/s11412-023-09409-w

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319-340. https://doi.org/10.2307/249008

Jishnu, D., Srinivasan, M., Dhanunjay, D. S., & Shamala, R. (2023). Unveiling student motivations: A study of ChatGPT usage in education. Shodh Kosh: *Journal of Visual and Performing Arts, 4*(2), 65-73. http://dx.doi.org/10.29121/shodhkosh.v4.i2.2023.503

Kayali, B., Yavuz, M., Balat, Ş., & Çalişan, M. (2023). Investigation of student experiences with ChatGPT-supported online learning applications in higher education. *Australasian Journal of Educational Technology, 39*(5), 20-39. https://doi.org/10.14742/ajet.8915

Khor, E. T., & Chan, L. (2025, June). Exploring the effect of scaffolding strategies in GenAI chatbot on student engagement and programming skill development. In *The Global Chinese Conference on Computers in Education (GCCCE) Proceedings* (pp. 32-39).

Khor, E. T., Chan, L., Koh, E., & Seow, P. (2026). Exploring students' perceptions of generative AI-generated feedback in learning programming. *Educational Psychology, 46*(1), 150–182. https://doi.org/10.1080/01443410.2025.2564894

Khor, E. T., Chan, L., & Seow, P. (2025). Students' use and perception of educational GenAI chatbot in high school computing: Insights from the decomposed theory of planned behavior. *Education Sciences, 15*(12), 1569. https://doi.org/10.3390/educsci15121569

Khor, E. T., Koh, E., & Seow, P. (2024). Leveraging Generative Artificial Intelligence to Support Students in Learning Programming: A Chatbot Pilot Study. In *Proceedings of the 18th International Conference of the Learning Sciences - ICLS 2024* (pp. 2217-2218).

Naeem, M., Ozuem, W., Howell, K., & Ranfagni, S. (2023). A step-by-step process of thematic analysis to develop a conceptual model in qualitative research. *International Journal of Qualitative Methods, 22*, 1-18. https://doi.org/10.1177/16094069231205789

Shoufan, A. (2023). Exploring students' perceptions of ChatGPT: Thematic analysis and follow-up survey. *IEEE Access, 11*, 38805-38818. https://doi.org/10.1109/ACCESS.2023.3268224

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science, 46*(2), 186-204. http://dx.doi.org/10.1287/mnsc.46.2.186.11926

Vo, A., & Nguyen, H. (2024). Generative artificial intelligence and ChatGPT in language learning: EFL students' perceptions of technology acceptance. *Journal of University Teaching & Learning Practice, 21*(6). http://dx.doi.org/10.53761/fr1rkj58

Vygotsky, L. S. (1978). Mind in Society: The development of higher psychological processes. Cambridge, MA: Harvard University Press.

Weisz, J. D., Muller, M., Ross, S. I., Martinez, F., Houde, S., Agarwal. M., Talamadupula, K., & Richards, J. T. (2022). *Better together? An evaluation of AI-supported code translation*. IUI '22: 27th International Conference on Intelligent User Interfaces, 369 - 391. https://doi.org/10.1145/3490099.3511157

Wu, T. -T., Lee, H. -Y., Li, P. -H., Huang, C. -N., & Huang, Y. -M. (2023). Promoting self-regulation progress and knowledge construction in blended learning via ChatGPT-based learning aid. *Journal of Educational Computing Research, 61*(8), 3 - 31. https://doi.org/10.1177/07356331231191125

Yeung, J. A., Kraljevic, Z., Luintel, A., Balston, A., Idowu, E., Dobson, R. J., Teo, J. T. (2023). AI chatbots not yet ready for clinical use. *Frontiers of Digital Health, 5*, Article 1161098. https://doi.org/10.3389/fdgth.2023.1161098

Zastudil, C., Rogalska, M., Kapp, C., Vaughn, J., & MacNeil, S. (2023). Generative AI in computing education: Perspectives of students and instructors. In 2023 IEEE Frontiers in Education Conference (FIE), Institute of Electrical and Electronic Engineers. https://doi.org/10.1109/FIE58773.2023.10343467

Zhang, Z., Dong, Z., Shi, Y., Price, T., Matsuda, N., & Xu, D. (2024). Students' perceptions and preferences of generative artificial intelligence feedback for programming. *Proceedings of the AAAI Conference on Artificial Intelligence, 38*(21), 23250-23258. https://doi.org/10.1609/aaai.v38i21.30372

Acknowledgments

This research is supported by the Ministry of Education (MOE), Singapore, under its Academic Research Fund (RG89/22). IRB reference no.: IRB-2023-392.

International Society of the Learning Sciences

Combining Collaborative and Adaptive Learning: Constructive–Interactive Engagement through Concept Maps, Teachable Agents, and an Adaptive Tutor

Yugo Hayashi, Ritsumeikan University, yhayashi@fc.ritsumei.ac.jp
Shigen Shimojo, Ritsumeikan University, sshimojo@fc.ritsumei.ac.jp
Kenneth R. Koedinger, Carnegie Mellon University, kk1u@andrew.cmu.edu

Abstract: Integrating intelligent tutoring systems (ITS) with computer-supported collaborative learning (CSCL) may support deeper learning by combining adaptive feedback with collaborative knowledge construction. This study examined whether integrating adaptive and collaborative supports promotes constructive–interactive engagement. Seventy-four university students participated in an experiment comparing three conditions: ITS-only (adaptive), concept map construction with a teachable agent (collaborative), and a hybrid condition combining both supports. Learning was measured using pre- and posttests, and collaborative processes were examined through concept maps and verbal interaction measures. Results showed that the hybrid condition achieved the highest learning gains, while adaptive support alone produced higher comprehension than collaborative support alone. Process measures showed no significant differences across conditions, suggesting that the hybrid condition maintained comparable levels of collaboration despite additional individual tasks. These findings suggest that integrating adaptive and collaborative supports can effectively support deeper learning in CSCL environments.

Introduction

Balancing individualized support and collaborative learning has become an important challenge in educational practice. Intelligent tutoring systems (ITS) provide adaptive feedback and mastery-based progression tailored to learners' understanding (Anderson et al., 1995; Koedinger & Aleven, 2007; VanLehn, 2011). Such adaptive support has proven effective for improving learning efficiency, for example through cognitive-model–based hint generation and mastery learning in systems such as the Open Learning Initiative(OLI) and Cognitive Tutor (Koedinger et al., 2013). However, ITS activities such as problem solving and response manipulation often remain at the *active* level of engagement, providing limited opportunities for learners to generate new interpretations or engage in reciprocal explanation (Graesser et al., 2014).

In contrast, computer-supported collaborative learning (CSCL) emphasizes peer interaction and knowledge co-construction through dialogue and shared artifacts (Roschelle & Teasley, 1995; Stahl et al., 2006). According to the ICAP framework (Chi & Wylie, 2014; Chi & Boucher, 2023), activities such as explanation, concept mapping, and collaborative reasoning correspond to *constructive* and *interactive* engagement, which are associated with deeper learning. Prior studies have shown that self-explanation facilitates deeper understanding (Chi et al., 1994), while peer explanation and questioning can promote reasoning and knowledge refinement (Miyake, 1986; Webb, 2009). Collaborative environments often incorporate externalization tools such as concept maps (Novak, 1990) and scripts or awareness systems that structure discussions and support knowledge building (Nesbit & Adesope, 2006; Fischer et al., 2013; Hayashi, 2020). Nevertheless, when learners differ in their level of understanding, some may dominate while others remain passive (Barron, 2003), and collaborative activities do not always lead to deep learning (Kirschner et al., 2006). These limitations partly stem from the limited capacity of CSCL environments to provide individualized adaptive support (Holmes et al., 2019; Roll & Winne, 2015).
Overall, ITS and CSCL exhibit complementary strengths and weaknesses. ITS provides individualized adaptive guidance but often elicits relatively shallow engagement, whereas CSCL promotes deeper constructive and interactive activities but lacks fine-grained adaptivity. Integrating these approaches may therefore support both individualized guidance and deeper collaborative engagement (Holmes et al., 2019; Roll & Winne, 2015).

This study examines whether combining adaptive support from ITS with collaborative activities supported by concept maps and a teachable agent can promote deeper learning. We conducted a controlled experiment comparing three conditions: (1) ITS-based adaptive learning, (2) collaborative learning using concept maps and a teachable agent (Hayashi et al., 2025), and (3) a hybrid environment integrating both supports. Learning outcomes were measured using pre–post comprehension tests, while collaborative processes were examined through concept maps and verbal interaction measures.

 © ISLS

Method

Participants

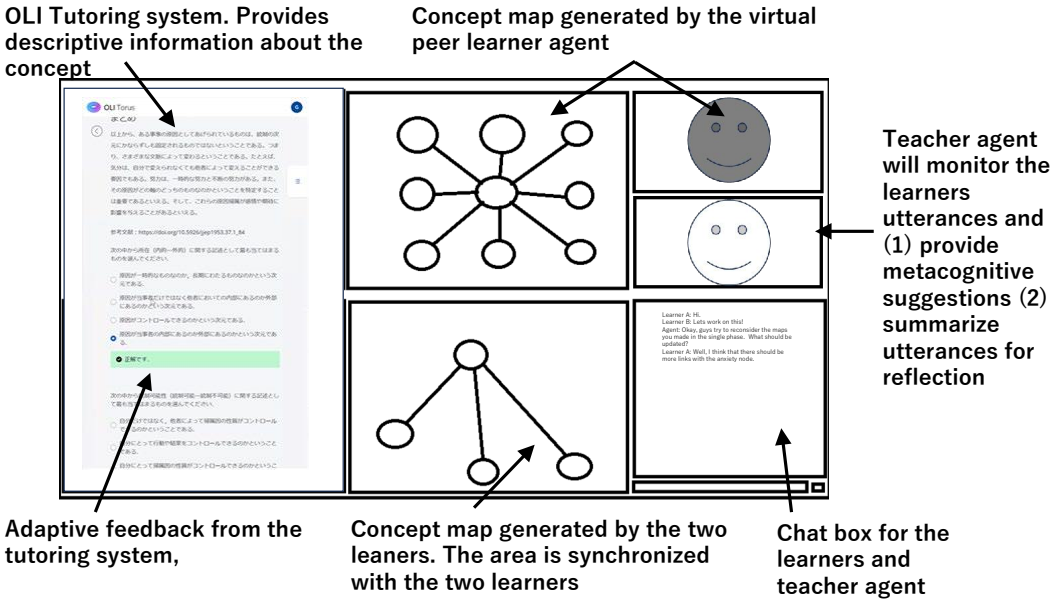
Seventy-four first- and second-year undergraduate students enrolled in an introductory psychology-related course at a comprehensive university in Japan participated in the study as part of a class activity. After receiving an explanation of the study and providing informed consent, participants were randomly assigned to one of three conditions: adaptive (n = 24), collaborative (n = 24), and hybrid (n = 26).

Learning environment

The learning environment combined the collaborative concept map system CoCot (Hayashi et al., 2025) with adaptive learning support from the OLI. The instructional materials addressed attribution theory in cognitive psychology (internal–external, stable–unstable, controllable–uncontrollable). The integrated system used in this study is illustrated in Figure 1.

Figure 1

Example of an Integrated Hybrid System



CoCot allows learners to construct and revise concept maps while interacting with peers and a teachable agent. Learners teach the agent using their concept maps and receive a generated map as feedback, encouraging constructive and interactive activities. However, CoCot alone does not provide individualized adaptive feedback. To provide adaptive guidance, the system was integrated with OLI, which delivers mastery-based learning support through immediate feedback and graduated hints based on learner responses (Koedinger et al., 2013). This integration enabled adaptive feedback within a collaborative learning environment.

Experimental design

Three conditions were implemented by activating different system functions:

1. Adaptive condition: learners worked individually using only OLI's adaptive support.
2. Collaborative condition: learners worked in pairs using CoCot to construct concept maps and interact with the teachable agent without adaptive feedback.
3. Hybrid condition: learners worked in pairs using both CoCot and OLI, combining collaborative interaction with adaptive feedback.

Experimental procedure

The experiment was conducted in a university computer laboratory. Participants first completed a tutorial and a pretest measuring prior knowledge. They then engaged in a 20-minute learning activity according to their assigned condition, followed by an individual posttest with the same format as the pretest. During the collaborative

conditions, all verbal interactions between learners were recorded by the system, and concept maps created during the session were stored for analysis.

Dependent variable

Learning outcomes were measured using pre–post comprehension tests scored on a five-point rubric (Hayashi, 2020). Two coders independently evaluated responses (Krippendorff's $\alpha = 0.81$). Collaborative processes were analyzed using two indicators. First, concept map construction activities were categorized into six operations: Create Node, Update Node, Delete Node, Create Link, Update Link, and Delete Link. Second, the total number of verbal utterances during collaborative activities was used as a quantitative indicator of interaction.

Results

Learning performance

A one-way ANOVA was conducted using the difference between pretest and posttest scores as the dependent variable. A significant difference was observed among conditions, ($F(2,71) = 26.96$, $p = .001$, $\eta_p^2 = 0.43$). Post hoc Tukey comparisons indicated that learning gains in the hybrid condition were significantly higher than in both the adaptive and collaborative conditions ($p = .001$ for both comparisons). In addition, the adaptive condition showed significantly higher learning gains than the collaborative condition ($p = .019$).

Concept mapping

Concept map construction was analyzed by categorizing operations into six types (Create Node, Update Node, Delete Node, Create Link, Update Link, and Delete Link). Table 1 shows the mean frequency of each operation for the collaborative and hybrid conditions. ANOVA results indicated no statistically significant differences between conditions (*ps* > .10). However, the hybrid condition showed a tendency toward more frequent node and link updates, suggesting comparable levels of knowledge externalization across conditions.

Table 1

Average Scores for the Constructive Concept Mapping Activity by Type

	CN	UN	DN	CL	UL	DL
Collaboration	6.29	58	0	6.45	6.29	0
condition	(4.50)	(47.48)	(0)	(4.42)	(5.16)	(0)
Hybrid	6.26	49.76	0.30	7.07	6.5	0
condition	(5.03)	(39.02)	(1.22)	(7.87)	(5.20)	(0)

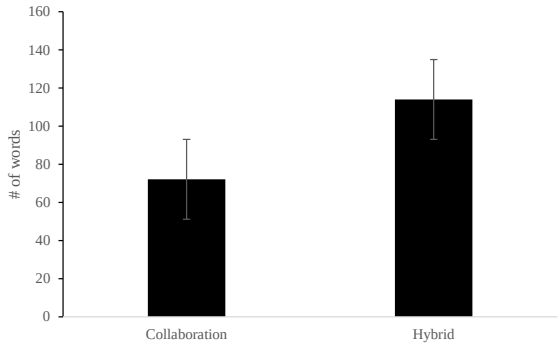
Figures in parentheses are standard deviations.

Utterances

Verbal interaction during collaborative activities was analyzed by counting the number of utterances produced by learners. Figure 2 shows the average number of utterances in the collaborative and hybrid conditions. No statistically significant difference was found ($p = .08$), although the hybrid condition showed a slightly higher mean number of utterances. This result suggests that collaborative dialogue was maintained in the hybrid condition despite the additional adaptive problem-solving activities.

Figure 2

Number of Utterances



Error bars represent standard deviations.

 © ISLS

Discussion

This study demonstrated the theoretical potential and practical significance of integrating ITS and CSCL to enhance learning outcomes. The hybrid learning environment that combined the OLI with collaborative concept mapping produced the highest comprehension gains, significantly outperforming both adaptive support alone and collaborative learning support alone. These findings suggest that individualized adaptive support stabilizes learners' foundational understanding of the content, which can then be further refined through collaborative activities. In addition, the higher performance of the adaptive condition compared with the collaboration-only condition indicates that adaptive guidance can effectively support initial knowledge acquisition before collaborative knowledge construction. Although the sequencing of adaptive and collaborative activities was not explicitly controlled in the hybrid condition, future research should examine how different sequencing strategies may further enhance learning outcomes.

Analyses of concept map operations and verbal utterances revealed no statistically significant differences between the collaborative and hybrid conditions, although the hybrid condition showed a slight tendency toward higher concept map activity. Importantly, even though learners in the hybrid condition engaged in additional adaptive problem-solving tasks, their concept map activity and verbal interaction levels remained comparable to those in the collaboration-only condition. This suggests that the added adaptive support did not reduce collaborative engagement and may have supported productive interaction. Future work should therefore examine collaborative processes over longer learning sessions and investigate qualitative aspects of dialogue and knowledge construction.

These findings also have implications for educational practice. Simply introducing collaborative learning activities may not always produce strong learning effects; however, combining collaboration with individualized adaptive feedback may enhance learning outcomes. The higher learning gains observed in the hybrid condition suggest that adaptive feedback may help learners correct misconceptions and better participate in collaborative knowledge construction. Such integrated support may be particularly useful in learning contexts where learners' levels of prior knowledge vary widely, such as first-year university education and large-scale online courses.

Nevertheless, several limitations should be noted. First, the measures of collaboration used in this study—concept map operations and utterance counts—capture only limited aspects of collaborative activity and do not directly reflect the quality of learners' interactions. Second, multiple instructional supports varied across experimental conditions, which limits the ability to isolate specific causal mechanisms underlying the observed effects. Third, the participants were limited to Japanese university students in a short-term experimental setting, and the generalizability of the findings to other contexts remains uncertain.

Future research should therefore investigate the qualitative features of collaborative interaction and their relationship to individualized adaptive support. In particular, discourse analysis grounded in the ICAP framework (Chi & Wylie, 2014) could help clarify how learners engage in active, constructive, and interactive activities during hybrid learning processes. Combining such analyses with learning analytics and AI-based modeling may further enable the design of real-time support that adapts to learners' collaborative processes.

Conclusion

Overall, the integrated condition was the most effective in improving learning comprehension, demonstrating the complementary roles of individualized adaptive support and collaborative interaction. Rather than treating ITS and CSCL as separate approaches, this study suggests that their integration can provide a productive framework for supporting deeper learning.

By combining adaptive feedback from ITS with knowledge externalization and interaction in CSCL environments, this study highlights how hybrid learning environments can support deeper understanding. These findings offer guidance for the design of next-generation learning environments that coordinate individualized guidance with collaborative knowledge construction.

Future studies should continue to investigate how hybrid environments influence learners' activity levels and collaborative processes, particularly through qualitative analyses grounded in the ICAP framework. Such work may contribute to bridging the ITS and CSCL research traditions and advancing integrated learning-support frameworks.

References

Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *The ournal of the Learning Sciences, 4*(2), 167–207.

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13*(2), 145–182.

Chi, M. T. H., & Boucher, A. R. (2023). The ICAP framework updated: Revisions and expansions. *Educational Psychologist, 58*(2), 120–138.
Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist, 49*(4), 219–243.
Dillenbourg, P. (2002). Over-scripting CSCL: The risks of blending collaborative learning with instructional design. In P. A. Kirschner (Ed.), *Three worlds of CSCL: Can we support CSCL?* (pp. 61–91). Open Universiteit Nederland.
Fischer, F., Kollar, I., Stegmann, K., & Wecker, C. (2013). Toward a script theory of guidance in computer-supported collaborative learning. *Educational Psychologist, 48*(1), 56–66.
Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2014). Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher, 43*(2), 90–102.
Hayashi, Y. (2020). Gaze awareness and metacognitive suggestions by a pedagogical conversational agent: An experimental investigation on interventions to support collaborative learning process and performance. *International Journal of Computer-Supported Collaborative Learning*, *15*(4), 469–498.
Hayashi, Y. Shimojo, S. Kawamura, T. (2025). Scripted Interventions versus Reciprocal Teaching in Collaborative Learning: A Comparison of Pedagogical and Teachable Agents Using a Cognitive Architecture. *Learning and Instruction*, *96*, 102057.
Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning*. Center for Curriculum Redesign.
Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*(2), 75–86.
Koedinger, K. R., & Aleven, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review, 19*(3), 239–264.
Koedinger, K. R., Booth, J. L., & Klahr, D. (2013). Instructional complexity and the science to constrain it. *Science, 342*(6161), 935–937.
Miyake, N. (1986). Constructive interaction and the iterative process of understanding. *Cognitive Science, 10*(2), 151–177.
Nesbit, J. C., & Adesope, O. O. (2006). Learning with concept and knowledge maps: A meta-analysis. *Review of Educational Research, 76*(3), 413–448.
Novak, J. D. (1990). Concept maps and Vee diagrams: Two metacognitive tools to facilitate meaningful learning. *Instructional Science, 19*(1), 29–52.
Roll, I., & Winne, P. H. (2015). Understanding, evaluating, and supporting self-regulated learning using learning analytics. *Journal of Learning Analytics, 2*(1), 7–12.
Roschelle, J., & Teasley, S. D. (1995). The construction of shared knowledge in collaborative problem solving. In C. O'Malley (Ed.), *Computer supported collaborative learning* (pp. 69–97). Springer.
Stahl, G., Koschmann, T., & Suthers, D. (2006). Computer-supported collaborative learning: An historical perspective. In R. K. Sawyer (Ed.), *Cambridge handbook of the learning sciences* (pp. 409–426). Cambridge University Press.
VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist, 46*(4), 197–221.
Webb, N. M. (2009). The teacher's role in promoting collaborative dialogue in the classroom. *British Journal of Educational Psychology, 79*(1), 1–28.

Acknowledgments

This study was supported by the JSPS KAKENHI (grant numbers JP23K28200, 23H03510 and 24KK0246).

International Society of the Learning Sciences

When Ideas Take Turns: Temporal Embodiment of Disciplinary Engagement

Xiaotian Zou, Daeun Hong, Cindy Hmelo-Silver, Chen Feng
xz107@iu.edu, dh37@iu.edu, chmelosi@indiana.edu, carrfeng@iu.edu
Indiana University Bloomington

Krista Glazewski, Tianshu Wang, James Lester
kdglazew@ncsu.edu, twang43@ncsu.edu, lester@ncsu.edu
North Carolina State University

Abstract: Collaborative problem-solving is central to science education, yet engaging disciplinarily in these settings is fluid, shifting across turns of talk in ways that shape group learning. Capturing these dynamics can provide insights for designing supports that foster deeper disciplinary engagement. This study examines how disciplinary engagement unfolds over time in a game-based science learning environment. Drawing on the Productive Disciplinary Engagement framework, we analyzed 2,415 discourse turns from 24 middle school groups using sequential and temporal analyses. Results revealed distinct pathways and temporal patterns across different groups, shedding lights into differences in learning outcomes. Findings demonstrate the value of temporal-sequential methods for informing scaffold design that supports productive collaboration.

Introduction and related work

Collaborative problem-based learning environments play a central role in science education by engaging students in authentic, complex inquiry (Dewi et al., 2023). However, group work alone does not guarantee meaningful disciplinary participation or learning success (Koretsky et al., 2021), prompting growing attention to how learners engage deeply with disciplinary content during collaboration (Fredricks, 2023).

A key lens for examining such engagement is Productive Disciplinary Engagement (PDE) (Engle & Conant, 2002), which describes learners' sustained participation in disciplinary practices as they co-construct knowledge and make collective intellectual progress (Engle, 2012; Cheng et al., 2019). PDE highlights four principles supporting productive collaboration: intellectual authority and responsibility, accountability to disciplinary norms and peers, problematizing through critique and questioning, and the use of resources to support reasoning. Because collaborative cognition is socially enacted through discourse (Jeong & Hmelo-Silver, 2016), disciplinary engagement is best understood as a dynamic interactional process.

Prior research has commonly used frequency-based coding to relate discourse features to learning outcomes (Yang et al., 2022), showing that successful groups engage more in idea generation and shared understanding (Popov et al., 2018). Yet count-based approaches may overlook the evolving and distributed nature of collaborative thinking (Lefstein et al., 2015). Recent work therefore emphasizes temporal and sequence-oriented analyses, demonstrating that ordered interaction patterns better predict learning outcomes than isolated discourse frequencies (Chen et al., 2018). Despite these advances, temporal approaches have rarely been applied specifically to the mechanisms of productive disciplinary engagement and associations with learning outcomes.

Addressing this gap, the present study integrates the PDE framework with sequence- and time-based analyses to operationalize engagement as a dynamic process. By examining transitions among PDE principles and comparing trajectories across groups with different performance levels, the study aims to identify mechanisms underlying productive collaboration and inform adaptive learning supports. To guide the analysis, we ask: what sequential patterns of PDE principles emerge, and how do they relate to group learning differences? how do temporal patterns of PDE principles unfold over time across groups with different learning outcomes?

Methods

Learning context, participants, and data sources

The study was conducted in *EcoJourneys* (Zou et al., 2024), a game-based computer-supported collaborative learning (CSCL) environment in which small groups of middle school students investigate the algae-related causes of illness affecting tilapia at a fish farm in the Philippines. The game includes four quests built around authentic scientific problems, in which students alternate between individual inquiry and collaborative activities *Deduce*

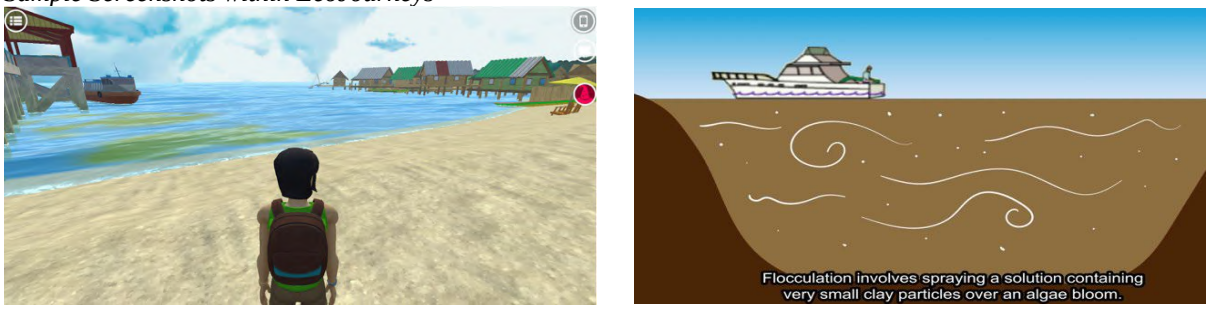
 © ISLS

and *TIDE*. This study focuses on Quest Four, a jigsaw-style activity emphasizing scientific argumentation and solution generation. Each student investigates two of four algae-reduction solutions (flocculation, aerators, filter feeders, and harvesters) through videos and background materials before returning to their group as an "expert." Groups collaboratively compare solutions through chat and face-to-face discussion, submit written justifications in *Deduce*, and develop persuasive proposals in *TIDE*. Responses are scored on a 1–3 scale for reasoning quality, clarity, and evidence use; combined Deduce and *TIDE* scores serve as indicators of group learning outcome. Figure 1 shows example screenshots from *EcoJourneys*, including individual inquiry (left) and a Quest Four informational video on flocculation (right).

The implementation took place in two middle schools in the Southeastern and Northeastern United States over 8–9 school days. From 31 recorded groups, 24 groups of four students were selected based on video quality. Groups were categorized as High-Performing (HP) (n = 11) or Low-Performing (LP) (n = 13) based on combined scores. Video-recorded face-to-face discourse was transcribed and synchronized with in-game chat using timestamp alignment, yielding 2415 discourse turns (1347 HP; 1068 LP), primarily from face-to-face interaction (1653 turns) with additional chat contributions (762 turns).

Figure 1

Sample Screenshots within EcoJourneys



Discourse coding and analytical approaches

We used an open coding approach to identify epistemic moves in students' discourse. Iterative coding produced 18 initial moves, which were consolidated into 16 codes capturing key collaborative problem-solving functions. Codes were mapped to the four Productive Disciplinary Engagement (PDE) principles—problematizing, authority, accountability, and resources (Engle & Conant, 2002). Three codes representing social coordination or non-disciplinary interaction were grouped under an additional Socialize category (Table 1). Two trained coders independently coded 20% of the data, achieving 83.29% agreement and Cohen's κ = .813. After resolving disagreements, the remaining data were coded by one coder.

Table 1

Disciplinary Engagement Coding Scheme

Main Categories	Epistemic Codes
Problematizing (PR)	1. Challenge: Directly or indirectly challenging the current approach
	2. Uncertainty: Statements that indicate confusion or a lack of information
	3. Limitation: Point out limitation or cons of a solutions
Authority (AU)	4. Position: Expressing agreement, preference, or disciplinary opinion
	5. Add: Expanding on prior contributions.
	6. Synthesize: Summarizing group thinking, signaling group summarization
	7. Evaluate: Making judgments about effectiveness, quality, of ideas, solutions.
Accountability (AC)	8. Probe: Requesting more information about a case or asking for clarification
	9. Solicit: Asking for group opinions or inviting group decisions
Resources (RE)	10. Elaborate: Providing a disciplinary explanation of how a case works.
	11. Cite: Citing directly the sources by referencing such as game characters or quotes
	12. Connect: Connecting to personal experience or metaphor
	13. Justify: Supporting a position or evaluation with reasoning, logic, or evidence.
Socialize (SO) (additional code added)	14. Coordinate: Organizing or managing the task or game-related actions
	15. Emotion: Statements expressing positive and negative emotions
	16. Sidechat: Statements that are non-disciplinary

We first conducted exploratory count-based analyses (Chen et al., 2018), examining contribution percentages, total turns, and average turn length across HP and LP groups using independent t-tests. As no

 © ISLS

significant differences were found, subsequent analyses focused on sequential and temporal discourse patterns. To address RQ1, we applied Lag Sequential Analysis to compute transitional probabilities among PDE principles, examining both direct transitions (lag = 1) and delayed transitions (lag = 2). Group differences were tested using two-way independent t-tests ($\alpha = .05$). We further conducted Frequent Sequence Mining to identify recurring multi-event discourse chains, analyzing three- and four-event sequences with a 3% support threshold. Analyses were implemented using the LagSequential and arulesSequences R packages. To address RQ2, discourse timelines were normalized by dividing each group's turns into 24 equal bins, aggregated into Early, Middle, and Late phases. Proportions of each PDE principle were calculated within each phase, and a series of two-way ANOVAs tested the effects of Time Phase and Group Type, as well as their interaction ($\alpha = .05$). Visualizations were produced using tidyverse and TraMineR.

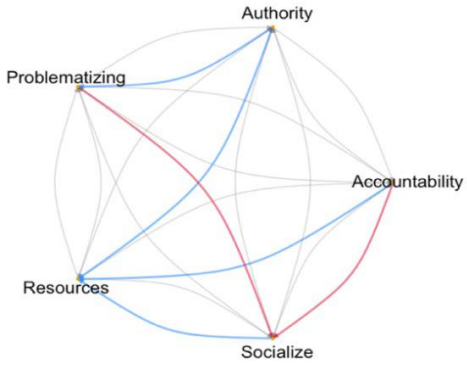
Results and discussion

RQ1: Sequential patterns among PDE principles by group types

Six transition pairs showed significant differences in transitional probabilities between HP and LP groups (Figure 3). Most significant transitions in HP groups shifted toward Resources as the subsequent move, suggesting that students frequently followed relational or decision-oriented discourse with evidence use, justification, or idea elaboration. The transition Authority → Problematizing ($p < .05$, $d = 0.34$) was also more likely in HP groups, indicating that expressed ideas often prompted critique and deeper epistemic engagement. In contrast, significant LP transitions more frequently led to Socialize, suggesting greater reliance on socioemotional or off-task interaction. Sequences such as Problematizing → Socialize and Accountability → Socialize indicate that critique or idea solicitation often shifted away from disciplinary inquiry. Lag-2 analysis revealed that HP groups were significantly more likely to exhibit indirect transitions from Problematizing → Resources ($p < .05$, $d = 0.29$), reinforcing the pattern that disciplinary resource use followed inquiry moves either immediately or after an intervening contribution. Frequent Sequence Mining (Table 2) further identified longer discourse patterns differentiating groups. HP groups showed sequences reflecting iterative cycles of critique, accountability, and authority (e.g., AC → PR → AC → AU), suggesting sustained collaborative reasoning. LP groups displayed more repetitive problematizing patterns with limited integration (e.g., PR → PR → RE), indicating potential confusion or weaker idea development despite small effect sizes for multi-event sequences.

Figure 3

Direct Transitional Probability by Group



Notes. Colored arrows show significant PDE transitions (clockwise). Blue indicates higher probabilities in HP Group; red indicates higher probabilities in LP Group.

Table 2

Top Frequent Multi-Event PDE Transitions by High- and Low-Performing Groups.

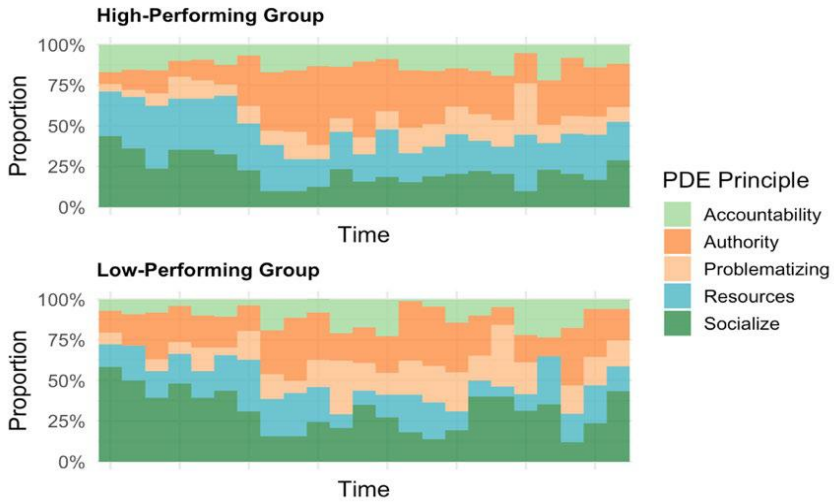
Higher frequency	Multi-event Sequence	Support			
		HP Group	LP Group	Cohen's *d*	*p*
High-Performing	AC→RE→RE	0.05	0.01	0.345	.028
	PR→AC→RE	0.04	0.01	0.316	.031
	AC→PR→AC→AU	0.03	0.00	0.287	.037
Low-Performing	PR→PR→RE	0.02	0.05	-0.341	.035
	PR→RE→AU→PR	0.00	0.03	-0.249	.042

 © ISLS

RQ2: Temporal patterns of PDE principles utilization by groups

Results from the two-way ANOVAs revealed differences in how High-Performing (HP) and Low-Performing (LP) groups employed PDE principles over the course of Quest Four (Figure 4). For Problematizing, a significant main effect of Time ($p = .010$, partial $\eta^2 = .56$) indicated increased use across phases, with a stronger rise in HP groups. This pattern suggests that HP groups engaged in critical examination later in the task, possibly after case presentations enabled comparison and disagreement. For Authority, both the main effect of Time Phase ($p = .053$) and the interaction ($p = .024$, partial $\eta^2 = .30$) were significant. HP groups showed increased authority use during the Middle phase, corresponding to the transition from sharing information to evaluating ideas, suggesting growing ownership and leadership in argumentation. LP groups peaked at Early and Middle stages but declined in the Late phase, indicating difficulty sustaining an authoritative stance during consensus building. No significant effects were found for Accountability, although descriptive increases appeared from Early to Middle phases in both groups. For Resources, a significant Group Type effect ($p < .001$, partial $\eta^2 = .28$) showed that HP groups consistently referenced evidence across all phases, particularly early in the task, suggesting stronger disciplinary grounding when group members shared cases with others. In contrast, LP groups relied less on resources and more on early authority statements, often proposing solutions before all information had been shared, as reflected in qualitative video observations. Finally, Socialize showed significant effects of Time ($p < .001$, partial $\eta^2 = .62$) and Group Type ($p = .009$, partial $\eta^2 = .28$). LP groups began with higher levels of socioemotional interaction, followed by a mid-task decline and slight late rebound, suggesting challenges transitioning from informal interaction to sustained task-focused collaboration.

Figure 4
Temporal Pattern of PDE Use by Groups



Conclusions

This study used sequential and temporal analyses to examine how disciplinary engagement unfolds during collaborative problem solving using PDE framework, revealing patterns not captured by traditional count-based measures. Several limitations should be noted. First, post-hoc video analysis showed that statistically identified transitions varied in timing and interactional quality; rapid exchanges sometimes reflected quick rebuttals rather than sustained engagement, suggesting that temporal spacing and interactional dynamics are not fully captured by sequential statistics (Wise & Padmanabhan, 2009). Second, mapping discourse to PDE principles required interpretation, and some interactions did not align cleanly with the framework. Although a Socialize category was added to capture socioemotional and coordination talk, this addition was post hoc and may limit generalizability. Despite these limitations, the study demonstrates the value of process-oriented analyses for understanding collaborative engagement. While count-based measures showed no group differences, sequential and temporal analyses revealed meaningful variations in how engagement developed over time, highlighting the importance of examining when and how discourse moves occur. Integrating quantitative analyses with qualitative video interpretation provided a more comprehensive account of engagement processes and identified transition patterns that may serve as indicators of productive collaboration and inform the design of scaffolds and learning analytics.

Future work should examine whether discourse sequences predict learning outcomes, support real-time detection of unproductive interaction patterns, and extend analyses across tasks and multimodal data sources to better understand how disciplinary engagement develops and is sustained in collaborative learning.

References

Chen, B., Resendes, M., Chai, C. S., & Hong, H. Y. (2018). Two tales of time: uncovering the significance of sequential patterns among contribution types in knowledge-building. *Learning Analytics, 20.*

Cheng, B., Gresalfi, M., Bell, A., Brady, C., Vogelstein, L., Damsa, C., ... & Ludvigsen, S. (2019). Theorizing and Measuring Collective Productive Disciplinary Engagement. *13th International Conference on Computer Supported Collaborative Learning (CSCL) 2019, Volume 2* (pp. 775-782)

Dewi, W. S., Siregar, R., Putra, A., & Hidayati, H. (2023). The effect of problem-based learning model on students' physics problem solving ability: A meta-analysis. *Jurnal Penelitian Pendidikan IPA, 9*(4), 2103-2109.

Engle, R. A. (2012). The productive disciplinary engagement framework: Origins, key concepts, and developments. *Design research on learning and thinking in educational settings* (pp. 161-200). Routledge.

Engle, R. A., & Conant, F. R. (2002). Guiding principles for fostering productive disciplinary engagement: Explaining an emergent argument in a community of learners classroom. *Cognition and Instruction, 20*(4), 399–483.

Zou, X., Hong, D., Chen, F., Hmelo-Silver, C. E., Glazewski, K., Wang, T., ... & Lester, J. (2024). Examine Student Resource Uses in a Game-based CSCL. In *Proceedings of the 17th International Conference on Computer-Supported Collaborative Learning-CSCL 2024, pp. 185-188*. International Society of the Learning Sciences.

Fredricks, J. A. (2023). Getting Students Engaged in Learning. State Education Standard, 23(3), 6-12. Healthcare, and Higher Education (pp. 3253-3261). Vancouver, BC: Association for the Advancement of Computing

Freedman, E. B., Hickey, D. T., Schamberger, B., & Harris, T. (2025). Reimaging productive disciplinary engagement and expansive framing: A synthesis of equity-oriented scholarship. *Educational Psychologist, 60*(2), 106-125.

Jeong, H., & Hmelo-Silver, C. E. (2016). Seven affordances of computer-supported collaborative learning: How to support collaborative learning? How can technologies help? *Educational Psychologist, 51*(2), 247-265.

Koretsky, M. D., Vauras, M., Jones, C., Iiskala, T., & Volet, S. (2021). Productive disciplinary engagement in high-and low-outcome student groups: Observations from three collaborative science learning contexts. Research in Science Education, 51(Suppl 1), 159-182. learning conversations. *Proceedings of E-Learn World Conference on E-Learning in Corporate, Government,*

Lefstein, A., Snell, J., & Israeli, M. (2015). From moves to sequences: Expanding the unit of analysis in the study of classroom discourse. *British Educational Research Journal, 41*(5), 866-885.

Li, T., Adah Miller, E., Simani, M. C., & Krajcik, J. (2025). Adapting scientific modeling practice for promoting elementary students' productive disciplinary engagement. *International Journal of Science Education, 47*(10), 1217-1251.

Olsen, J. K., Sharma, K., Rummel, N., & Aleven, V. (2020). Temporal analysis of multimodal data to predict collaborative learning outcomes. British Journal of Educational Technology, 51(5), 1527-1547.

Popov, V., Xing, W., Zhu, G., Horwitz, P., & McIntyre, C. (2018). The influence of students' transformative and non-transformative contributions on their problem solving in collaborative inquiry learning. *2018 International Conference on Learning Sciences (ICLS)*

Wise, A., & Padmanabhan, P. (2009). Developing structural-temporal visualizations of asynchronous threaded online learning conversations. *In eLearn: World Conference on EdTech* (pp. 3253-3261)

Yang, Y., Yuan, K., Feng, X., Li, X., & van Aalst, J. (2022). Fostering low-achieving students' productive disciplinary engagement through knowledge-building inquiry and reflective assessment. *British Journal of Educational Technology, 53*(6), 1511-1529.

Zou, X., Hong, D., Chen, F., Hmelo-Silver, C. E., Glazewski, K., Wang, T., ... & Lester, J. (2024). Examine Student Resource Uses in a Game-based CSCL. In *Proceedings of the 17th International Conference on Computer-Supported Collaborative Learning-CSCL 2024, pp. 185-188*. International Society of the Learning Sciences.

Acknowledgement

This research was supported by funding from the National Science Foundation (NSF) under Grant DRL-2112635. Any opinions, findings, and conclusions expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

International Society of the Learning Sciences

Participatory Design for AI in Teacher Education: From Top-Down Directives to Teacher Empowerment

Ekta Shokeen, Aarhus University, ekta.shokeen@cc.au.dk
Ane Vielandt Jensen, Aarhus University, av@cc.au.dk,
Ole Sejer Iversen, Aarhus University, oiversen@cc.au.dk
Christian Dindler, Aarhus University, dindler@cavi.au.dk
Marianne Graves Petersen, Aarhus University, mgraves@cs.au.dk

Abstract: Empowering teacher educators to lead technology integration is essential for creating context-sensitive educational futures. This study examines how participatory design workshops can support educators in collaboratively developing sustainable capacity for integrating AI into teacher education. Drawing on a series of three participatory design workshops, the analysis focuses on the collaborative learning processes involved and the professional outcomes they produce. Structured around exploration, co-creation, and reflection, these workshops enabled educators to critically engage with the pedagogical, ethical, and institutional dimensions of AI, fostering both ownership and transformative agency. The findings demonstrate how participatory design can function not only as a design methodology but also as a framework for professional learning. A key contribution of this paper is the demonstration that participatory design supports sustainable AI integration by building educators' capacity for change through the design process.

Introduction

Artificial intelligence (AI) is reshaping education through adaptive learning systems, automated assessment, and data-driven decision-making (e.g., Holmes et al., 2019), raising urgent questions about both classroom practice and how teachers and teacher educators (TEs) are equipped to navigate this shift (Chounta et al., 2022; Lin & Van Brummelen, 2021). While AI promises innovation, its implementation often follows top-down models that marginalize TEs, reducing their role to passive adopters of pre-defined tools and policies (e.g., Zawacki-Richter et al., 2019; Williamson & Eynon, 2020).

This study argues that empowering TEs to actively shape technology is critical , not only for successful AI integration but for sustaining context-sensitive, pedagogically grounded change. Drawing on participatory design (PD) and Cultural-Historical Activity Theory (CHAT), we position TEs as co-creators rather than end-users, capable of engaging critically with AI to transform educational practices, policies, and resources (Ehn, 1993; Bødker et al., 2000; Sannino, 2015). Specifically, we examine:

How can participatory design workshops empower teacher educators to build capacity for integrating AI in teacher education?

Background

As digital technologies increasingly shape education, teacher empowerment has emerged as a central concern. The challenge extends beyond effective tool use to supporting educators in shaping the values, purposes, and pedagogical practices that guide technology integration. AI exemplifies this challenge: while it enables personalization, automated feedback, and data-informed instruction, it also raises ethical and pedagogical dilemmas related to bias, authorship, academic integrity, and the role of human judgment. TEs, who prepare future teachers, are therefore pivotal. Yet they are often positioned as passive recipients of AI policies and tools rather than as active agents shaping their implementation (Williamson & Eynon, 2020; Zawacki-Richter et al., 2019).

Emerging research challenges this passive framing. Evidence suggests that when teachers engage in participatory curriculum design, both contextual relevance and professional agency are strengthened (Lin & Van Brummelen, 2021; Tatar et al., 2024). Teachers recognize AI's instructional potential but consistently emphasize the need for ethical and pedagogical preparedness (Chounta et al., 2022). AI integration can disrupt evaluative roles, prompting reflection on authorship and academic integrity (Mohammadkarimi, 2023), highlighting the centrality of teachers' beliefs, values, and professional judgment in responsible adoption.

Teachers' attitudes toward AI are marked by both optimism and caution. While curiosity may drive engagement, meaningful and confident use depends on foundational AI literacy and ethical awareness (Long & Magerko, 2020; Edwards et al., 2018). Concerns about algorithmic bias, fairness, reduced student interaction, and

 © ISLS

overreliance on automation underscore the need for professional development approaches that go beyond technical training to cultivate critical, reflective, and empowered engagement (Kaplan-Rakowski et al., 2023).

Drawing on Iversen's (2019) perspectives on digital technology this study foregrounds democratic participation and critical agency of educators where AI is approached not merely as an instructional aid but as a complex phenomenon that requires ethical reflection and value-driven integration within teacher education.

Theoretical framing

PD offers a robust framework for exploring teacher empowerment in AI integration. Widely applied in educational contexts to develop curricula and learning environments (Könings et al., 2014), PD foregrounds teachers' professional expertise and lived experiences. Methods such as prototyping, storytelling, and scenario building foster mutual learning, make assumptions visible, and support iterative refinement of ideas (DiSalvo et al., 2017). Beyond methodology, PD embodies an ethical and democratic stance, promoting shared ownership and critical engagement with technology (Durall & Kapros, 2020; Cumbo & Selwyn, 2022). Although extensively applied in K–12 contexts, its use with teacher educators (TEs) remains underexplored, despite their key role in linking classroom innovation to systemic change. Involving educators in PD enhances adoption and confidence in integrating emerging technologies (Severance et al., 2018; Durall & Kapros, 2020).

This study positions PD as an approach to cultivate transformative agency, *the collective capacity to question practices, envision alternatives, and reshape educational systems*, grounded in Scandinavian traditions emphasizing empowerment, ethics, and democracy (Ehn, 1993; Bødker et al., 2000). To deepen PD methodological approach, we draw on CHAT, which, like PD, focuses on socially situated activity systems and the generative potential of contradictions (Engeström, 2014). AI integration often exposes tensions between institutional rules, pedagogical aims, and technological affordances, which CHAT frames as drivers of expansive learning (Engeström, 2014). By linking PD and CHAT, workshops become both practical design spaces and sites of empowerment, enabling TEs to develop the agency needed for ethically grounded, sustainable AI integration (Sannino, 2015; Virkkunen, 2006).

Methods

This study was conducted within the Danish research project, which investigates AI integration in Danish teacher education. Nine teacher educators (TE1–TE9) from four university colleges participated, representing first language, art, pedagogical psychology, didactics, and pedagogy. PD workshops served as both intervention and data-collection sites, engaging TEs as practitioner collaborators in co-developing AI-integrated lesson plans. The study received ethics approval from Aarhus University, and informed consent was obtained from all participants.

Table 1

Workshop Activities Schedule

Activity	Workshop 1	Workshop 2	Workshop 3
A1	Introduction	Introduction	Introduction
A2	Initial reflections on AI (P, 40 min)	Technical – How does AI work? (P, 45 min)	Perspectives on generative AI (P, 30 min)
A3	Perspectives on AI (P, 20 min)	Perspectives on AI and AI literacy (P, 20 min)	Reflections on teaching AI experiences (D, 75 min)
A4	Relevance of AI literacy (P, 20 min)	Reflections on AI integration (D, 15 min)	AI & computational empowerment (P, 45 min)
A5	Technical – How does AI work? (P, 15 min)	Preparing lesson plan (150 min)	Exploring further connections (D, 90 min)
A6	Hands-on with *The Engine Room* & ChatGPT (50 min)	-	-
A7	Reflections on AI integration (D, 50 min)	-	-
A8	Drafting lesson plan (15 min)	-	-

Note: 'P' = Presentation; 'D' = Discussion. Duration of activities shown in parentheses.

Three in-person workshops were held over seven months, facilitated by eight researchers from five institutions across three locations. Workshops 1–2 focused on mutual learning and co-creation through small and large-group activities; Workshop 3 emphasized post-implementation reflection. Each workshop followed iterative stages: framing, exploration (including AI misconceptions), ideation, prototyping, and reflection. Activities (refer to Table 1) included collaborative design tasks and hands-on use of AI tools.

Data included 11.5 hours of workshop audio, images, video clips, artifacts, and seven semi-structured interviews (31–52 minutes; M=40). Recordings were transcribed in Condens and verified. Thematic analysis guided coding using a hybrid deductive–inductive approach (Saldana, 2011). An initial codebook, based on the research question, guided coding while allowing emergent themes. Three data subsets were independently coded by two researchers to refine codes, which were then applied to the full dataset. Weekly team discussions resolved discrepancies, clarified code meanings, and ensured consensus, enhancing reliability and transparency. This approach enabled systematic, participant-centered identification of themes reflecting perspectives.

Findings

1.From individual uncertainty to collective direction

Across the PD workshops, collaboration functioned as the central mechanism through which TEs developed transformative agency for AI integration. Rather than merely exchanging lesson ideas, participants engaged in collective sense-making that surfaced shared goals, tensions, and new possibilities for teacher education.

From the outset, TEs articulated both personal and collective objectives. Some sought professional growth. TE7 admitted wanting "*to get a more nuanced view of it - right now, I don't think I have a particularly nuanced one, because I don't know enough about it*." Others framed AI integration as a systemic responsibility. TE5 argued that AI should be "distributed in such a way that it becomes everyone's responsibility as an educator in teacher education," while TE6 emphasized curricular progression so that "*pre-service teachers aren't just introduced to AI in the same way every time… but that there's also thought given to progression*." Through dialogue, these perspectives converged into a shared vision of cross-curricular, developmentally structured AI integration marking a shift from individual uncertainty toward collective direction.

The PD workshops fostered multiple forms of collaboration among TEs and facilitators, across three interrelated dimensions: *articulating objectives*, *negotiating attitudes*, and *addressing concerns*. These collaborations unfolded shared inquiry, co-construction of knowledge, and the cultivation of agency.

2. Negotiating tensions and ethical concerns

The workshops also created space to negotiate conflicting attitudes. AI was not approached with uniform enthusiasm. TE7 described the surrounding pressures as "*completely draining," and TE2 worried that, within first-language education, "there's an educational or formative perspective that risks being lost*." Yet these concerns were met with curiosity and openness. TE5 emphasized engagement "so it doesn't have to be rooted in fear or anxiety but can instead become something more positive."

Rather than fragmenting the group, the juxtaposition of skepticism and optimism became a resource for shared inquiry. Participants critically examined AI's purpose and scope. TE6 asked, "*Is it about the empowerment perspective? Is it about focusing on computational thinking and understanding of data and algorithms? Or where are we, really*?" TE8 questioned proportionality: "*To what extent should it take up space—are we supposed to spend every day from morning to night for the rest of our lives discussing AI?"*

Ethical dimensions were equally central to dialogue during PD. TE3 warned that "*what the majority says ends up becoming the truth, and the nuances disappear,*" while TE8 stressed that "*algorithms are in control—that's where computational empowerment comes in—it's partly about highlighting bias.*" Even vulnerability became generative; TE4 acknowledged, "*I am navigating this somewhat blindly*." Such admissions transformed uncertainty into shared capacity-building, reinforcing collaboration as a driver of transformative agency.

3. PD as capacity building and reframing AI

Hands-on PD activities grounded TEs reflections in practice. As TE3 described, the workshops offered "*a mini hook and hang something on it. And then you can unfold it yourself in your subject*." This flexible structure supported disciplinary adaptation. TE5 valued "*all the different perspectives,*" noting how collaborative exchanges enriched planning.

Different pedagogical entry points emerged. Some began with technical foundations: TE1 explained how AI relies on "*generating language through coding and a kind of probabilistic reasoning,*" and TE3 emphasized that AI "*is not creative; it is a machine… based on statistics and vast amounts of data*." Others foregrounded socio-cultural themes. TE5 connected AI to explorations of loneliness and socialization, while TE4 described how experimenting with School-GPT in writing tasks shifted AI from being "*both a tool, but also a subject area,something we need to know, be able to use, and understand*."

Across these trajectories, a common progression emerged during PD workshops, for TEs AI moved from being treated primarily as a tool to becoming a site of empowerment and critical reflection. TE7 described structuring learning, so pre-service teachers begin with experience and "*suddenly… you're able to go out and act*

on it." During PD dialogues TE6 framed AI as "*a catalyst for societal change and transformation*." These findings suggest PD potential to build capacity and reframing TEs perceptions of AI integration in education.

Discussion

This study examined how collaboration within PD workshops support TEs in developing transformative agency for AI integration. Drawing on CHAT, the findings show that PD workshops functioned not merely as training spaces but as collective arenas where contradictions in existing practices were surfaced, negotiated, and reworked. Rather than positioning AI as a top-down reform, the workshops enabled TEs to question prevailing assumptions, articulate shared purposes, and experiment with new pedagogical directions. In doing so, participants moved from adapting to technological change toward actively reshaping the conditions of their professional practice.

Consistent with AI literacy research emphasizing confidence, critical evaluation, and responsible adoption (Edwards et al., 2018; Chounta et al., 2022), the workshops created space for TEs to clarify goals, confront ethical dilemmas, and reconcile divergent attitudes. Dialogue in PD workshops supported a shift from uncertainty and skepticism toward shared, value-driven visions of AI integration. Through co-construction of lesson ideas and exchange of disciplinary expertise, participants developed ownership, adaptability, and confidence outcomes aligned with prior studies on collaborative professional learning (Severance et al., 2018; Durall & Kapros, 2020; Lin & Van Brummelen, 2021).

From a CHAT perspective, contradictions between optimism and skepticism, institutional mandates and pedagogical values, technical functionality and ethical responsibility, acted as drivers of expansive learning (Engeström, 2014; Sannino, 2015). PD workshops made these tensions visible and discussable. Participants reframed pressures surrounding AI as shared challenges rather than individual burdens, collectively interrogating the rules, roles, and tools that shape their activity systems. In this sense, contradiction became a generative resource for systemic reflection and redesign.

The identified collaborative practices articulating objectives, negotiating attitudes, and addressing concerns map onto dimensions of transformative agency (Virkkunen, 2006). By envisioning progression in teacher education, questioning dominant technological discourses, and foregrounding ethical implications, TEs enacted agency as a collective, emergent process rather than an individual trait. PD methods grounded this agency in practice. Hands-on experimentation with AI tools, prototyping, and structured reflection bridged abstract critique and concrete implementation. Diverse trajectories of integration technical, socio-emotional, or critical-analytical were accommodated, yet converged on empowering pre-service teachers to engage critically with AI. This flexibility demonstrates how PD can honor disciplinary differences while sustaining shared ownership. Importantly, participants reframed AI integration around democratic and emancipatory purposes. AI was positioned not simply as a tool to master but as a phenomenon to question, shape, and situate within broader societal aims. Such reframing reflects expansive learning: redefining the object of teacher education from preparing students to use AI toward preparing them to critically influence its role in society.

Overall, the study positions PD workshops as collaborative knowledge-building environments where professional learning, identity, and systemic innovation are co-constructed. For CSCL, it highlights democratic participation and collective design as central to cultivating TEs transformative agency in AI integration.

Limitations and next steps

These contributions should be considered in light of several limitations. Although participant diversity in expertise was intentional, the findings cannot be generalized to any single subject domain within teacher education. Additionally, the analysis emphasizes established dimensions of PD in teacher education rather than foregrounding innovative applications of AI itself, an important direction for future inquiry.

Conclusion

This study shows that empowering TEs to integrate AI requires more than tools or technical training; it requires approaches that cultivate agency, collaboration, and critical engagement. Framing PD as both a methodology and an ethical stance grounded in empowerment and democracy, the study demonstrates how PD workshops can build sustainable capacity for AI integration in teacher education. Through iterative cycles of exploration, co-creation, and reflection, TEs surfaced contextual tensions and reimagined practices aligned with their professional values. AI became a site for inquiry and collective learning rather than a top-down mandate. Methodologically, the study positions PD workshops as spaces for shared knowledge construction and highlights democratic participation as central to educator-led, sustainable digital transformation.

References

Bødker, S., Ehn, P., Sjögren, D., & Sundblad, Y. (2000). Co-operative design—perspectives on 20 years with "the Scandinavian IT Design Model". *Proceedings of NordiCHI*, 2000, 22–24.

Chounta, I.-A., Bardone, E., Raudsep, A., & Pedaste, M. (2022). Exploring teachers' perceptions of Artificial Intelligence as a tool to support their practice in Estonian K-12 education. *International Journal of Artificial Intelligence in Education*, 32(3), 725–755.

Connelly, L. (2023). The engine room. https://maskinrummet.github.io/#/en

Cumbo, B., & Selwyn, N. (2022). Using participatory design approaches in educational research. *International Journal of Research & Method in Education*, 45(1), 60–72.

DiSalvo, C., Clement, A., & Pipek, V. (2017). Participatory design for, with, and by communities. International handbook of participatory design, 182-209.

Durall, E., & Kapros, E. (2020). Co-design for a Competency Self-assessment Chatbot and Survey in Science Education. In *International conference on human-computer interaction* (pp. 13-24). Springer International Publishing.

Edwards, C., Edwards, A., Spence, P. R., & Lin, X. (2018). I, teacher: Using artificial intelligence (AI) and social robots in communication and instruction. *Communication Education*, 67(4), 473–480.

Ehn, P. (1993). Scandinavian design: On participation and skill. In D. Schuler & A. Namioka (Eds.), *Participatory Design* (pp. 41–77). CRC Press.

Engeström, Y. (2014). *Learning by expanding: An activity-theoretical approach to developmental research*. Cambridge University Press.

Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education promises and implications for teaching and learning.* Center for Curriculum Redesign.

Iversen, O. S., Dindler, C., & Smith, R. C. (2019). *A Design Approach to Understanding Technology*. Dafolo.

Kaplan-Rakowski, R., Grotewold, K., Hartwick, P., & Papin, K. (2023). Generative AI and teachers' perspectives on its implementation in education. *Journal of Interactive Learning Research*, 34(2), 313–338.

Könings, K. D., Seidel, T., & Van Merriënboer, J. J. G. (2014). Participatory design of learning environments: Integrating perspectives of students, teachers, and designers. *Instructional Science*, 42(1), 1–9.

Lin, P., & Van Brummelen, J. (2021). Engaging Teachers to Co-Design Integrated AI Curriculum for K-12 Classrooms. Proceedings of the 2021 *CHI Conference on Human Factors in Computing Systems*, 1–12.

Long, D., & Magerko, B. (2020). What is AI Literacy? Competencies and design considerations. Proceedings of the 2020 *CHI Conference on Human Factors in Computing Systems*, 1–16.

Mohammadkarimi, E. (2023). Teachers' reflections on academic dishonesty in EFL students' writings in the era of artificial intelligence. *Journal of Applied Learning and Teaching*, 6(2), 105–113.

Saldana, J. (2011). Fundamentals of qualitative research. Oxford university press.

Sannino, A. (2015). The emergence of transformative agency. *Learning, Culture and Social Interaction*, 4, 1–3.

Severance, S., Penuel, W. R., Sumner, T., & Leary, H. (2018). Organizing for teacher agency in curricular co-design. In *Cultural-historical activity theory approaches to design-based research* (pp. 45–78). Routledge.

Tatar, C., Jiang, S., Rosé, C. P., & Chao, J. (2024). Exploring Teachers' Views and Confidence in the Integration of an Artificial Intelligence Curriculum into Their Classrooms: A Case Study of Curricular Co-Design Program. *International Journal of Artificial Intelligence in Education*.

Virkkunen, J. (2006). Dilemmas in building shared transformative agency. Activités, 3(3-1).

Williamson, B., & Eynon, R. (2020). Historical threads and future directions in AI in education. *Learning, Media and Technology*, 45(3), 223–235.

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education–where are the educators?. *International journal of educational technology in higher education*, 16(1), 1-27

Acknowledgments

We sincerely thank the participating teachers for their invaluable contributions to this work. This research was supported by the Ministry of Higher Education and Science in Denmark (Grant No. 4301-00001B). The views expressed in this paper are those of the authors and do not necessarily reflect the position of the funding agency.

International Society of the Learning Sciences

DotQuest: Accessible STEM Games to Support Blind and Visually Impaired Learners

Ariel Han, Utah State University, ariel.han@usu.edu

Rajeev Joshi, Daniel Tsivkovski, Dylan Ravel, Maryam Etezadbrojerdi, Franceli Cibrian
rajoshi@chapman.edu, tsivkovski@chapman.edu, ravel@chapman.edu, etezadbrojerdi@chapman.edu, cibrian@chapman.edu
Chapman University

Abstract: Blind and low vision (BLV) students face persistent barriers to full participation in STEM learning, particularly due to the limited availability of accessible tools. This study introduces DotQuest, a technology probe that integrates haptic feedback, and a conversational agent to create narrative-driven, game-based STEM activities. To examine how such tools might support equitable participation, we conducted interviews with seven Teachers of Students with Visual Impairments (TVIs) across U.S. Analysis revealed that teachers emphasize the importance of simultaneous tactile and auditory access, the need for classroom tools that foster both independence and inclusion, and the value of designing accessibility from the outset rather than as an accommodation. Findings show that multimodal game-based environments can enhance STEM engagement while revealing systemic barriers for BLV learners. This work contributes to the learning sciences by deepening understanding of how multimodal and inclusive design can guide the development of equitable STEM learning environments.

Introduction

STEM subjects often rely heavily on visual representations, yet equivalent multimodal tools that integrate tactile, auditory, and interactive supports remain limited (Ávila-Soto et al., 2017). As a result, Blind and low vision (BLV) students are frequently excluded from the kinds of rich, hands-on learning experiences that promote conceptual understanding, motivation, and equitable engagement (Wu et al., 2025). This study is motivated by the goal of designing accessible, tangible, and interactive learning environments that expand STEM opportunities for BLV learners (Pires et al., 2019; Lang et al., 2023; Kim & Yeh, 2015).

This study represents the first phase of a design-based research (DBR) project aimed at iteratively designing and refining DotQuest, an accessible, multimodal, game-based STEM learning environment for BLV students (Anderson & Shattuck, 2012). To connect theoretical perspectives on learning with the designed features and observed processes of our intervention, we used conjecture mapping (Sandoval, 2014). This phase focuses on eliciting expert insights to refine the system's design conjectures, features, and pedagogical functions before classroom implementation (Sandoval, 2014). By engaging Teachers of Students with Visual Impairments (TVIs) as co-design informants, we seek to connect theoretical principles of multimodal and inclusive learning with the pragmatic realities of STEM teaching for BLV learners (Hutchins & Biswas, 2024).

This work contributes to the Learning Sciences by articulating and refining design conjectures that link multimodal accessibility, narrative game structures, and tactile–auditory integration to equitable STEM engagement for blind and low-vision learners (Chen et al., 2024). Through teacher-informed conjecture mapping, DotQuest demonstrates how inclusive, multimodal game design can operationalize theoretical principles of sociocultural learning (Vygotsky, 1978) and Universal Design for Learning (UDL) into tangible design features (Rose, 2000). The study contributes a model for bridging accessibility research and design-based inquiry, expanding how the field conceptualizes equity-oriented learning environments beyond visual modalities (Uttamchandani, 2018). This study was guided by the following research questions: (1) *How do TVIs conceptualize the design features and pedagogical affordances needed for multimodal, game-based STEM learning environments for blind and low-vision learners?* (2) *How do TVIs' insights inform the refinement of DotQuest's design conjectures that link tactile–auditory integration, narrative-based interaction, and inclusive design principles to equitable STEM engagement?*

 © ISLS

Background

Multimodal learning and accessibility for BLV students

The Learning Sciences community has explored how blind and low vision (BLV) students engage in learning. For instance, researchers such as Foley et al. (2025) examine strategies to strengthen tactile graphicacy for BLV learners, emphasizing the critical role of teachers of students with visual impairments (TVIs) in scaffolding tactile sensemaking to support nonvisual spatial literacy. More broadly, research in BLV education highlights the centrality of tactile and auditory modalities as primary channels for conveying information and underscores educators' role in scaffolding these multimodal learning processes (Foley et al., 2025; Kari et al., 2006).

This study is grounded in sociocultural theories of learning, constructivist, UDL, From a sociocultural perspective (Vygotsky, 1978), learning is mediated by social interaction and scaffolding highlight how tactile, game-based interaction (Plass et al., 2015), and auditory modalities can scaffold sensemaking in non-visual ways, allowing BLV students to engage in STEM content.

Methods

Participants

We conducted virtual interviews via Zoom with seven TVIs across the United States. Each individual interview lasted approximately 60 minutes. Participants were recruited through the researchers' professional and personal networks. Of the seven participants, six identified as female and one as male. Their professional experience ranged from a minimum of 11 years to over 30 years. All participants were employed in public school districts . We conducted teacher interviews as part of the first design research cycle to elicit expert feedback that would inform the refinement of DotQuest's design conjectures, features, and functions. By engaging Teachers of Students with TVIs as design partners, we sought to surface pedagogical insights that could shape the next iteration of the system.

Data sources and analysis

The Zoom interviews were recorded and transcribed using an automatic transcription tool, followed by manual review for accuracy. Each interview began with demographic questions, including teachers' years of experience and the age ranges they serve. We then asked general questions about challenges in their teaching practices during the first 15 minutes. Following this, we introduced a technology probe by presenting our prototype, *DotQuest*, and invited feedback on its applicability and affordances to elicit further design and pedagogical insights. This study was reviewed and approved by the research institute's Institutional Review Board (IRB).

For analysis, we adopted an inductive approach to generate initial codes and identify emergent patterns in the data. To ensure analytic rigor, the research team engaged in iterative coding sessions, meeting weekly over the course of a month to refine codes, resolve discrepancies, and reach consensus. This process emphasized collaborative meaning-making and reflexivity, strengthening the trustworthiness of our interpretations. The final set of codes was then organized into broader thematic categories, including: Simultaneous Tactile + Verbal Access for Multisensory Instruction, Accessible Game-Based Learning for Equitable Engagement, and Inclusive Design in Educational Technology & Classroom Orchestration.

DotQuest: Accessible game design for equitable STEM engagement

DotQuest is a technology probe designed to explore how multimodal, game-based environments can make STEM learning more accessible for BLV students. The system integrates three key components: (1) a refreshable Braille display (RBD) for tactile interaction and Braille literacy (Autonomy), (2) haptic feedback for conveying changes such as temperature, numbers, shapes, or movement (Inclusion), and (3) a conversational agent that guides learners through narrative-driven "quests. (Competence)" Each quest is framed around an interactive story, for example, purchasing ingredients for a recipe or exploring seasonal changes, allowing learners to experience core STEM concepts (e.g., counting, measurement, sequencing) through tactile and auditory modalities. DotQuest structures learning as a series of gamified "Quests" that embed STEM problem-solving within narrative and embodied interaction. Each Quest situates the learner in a story-driven context where the digital character, Dot, acts as a social and pedagogical partner; scaffolding, prompting reflection, and adapting to the learner's evolving understanding. This design leverages narrative framing and game mechanics (e.g., progression, feedback loops, and rewards) to sustain engagement while aligning with constructionist and sociocultural principles of learning through doing and dialogue.

 © ISLS

Findings

Simultaneous tactile + verbal access for multisensory instruction

Teachers highlighted that while existing tools offering auditory support are useful for practice, they fall short of providing the rich, multisensory experiences necessary for deeper learning. In particular, they emphasized that integrating Braille with auditory feedback could be transformative for accessibility, enabling BLV students to engage more meaningfully with STEM concepts. For instance, Ms. Cruz (25 years of experience) said: *"A device like that would provide both auditory and tactile feedback, which is critical for these students. Right now, many assistive technologies only offer one type of feedback, which isn't always enough. I'd love to see how this develops!"*

Teachers emphasized that auditory-only tools are helpful but insufficient. They highlighted the need for integrating tactile and auditory modalities, especially refreshable Braille displays with verbal feedback, as transformative for accessibility, deeper learning, and equitable STEM engagement for BLV students. This will support the *Inclusion* principle, as the game will provide equitable access to the game to accommodate multiple sensory modalities rather than just rely on vision or hearing.

Inclusive design in educational technology and classroom orchestration

Teachers emphasized that accessibility must be built into educational technologies from the outset rather than treated as an afterthought. They noted that assistive tools for BLV students often lag behind general education technologies, creating persistent inequities. As Ms. Wynn (14 years of experience) explained: *"If all educational software were designed with accessibility in mind from the start, it would make a huge difference. Technology also struggles to keep pace with advancements in general education."*

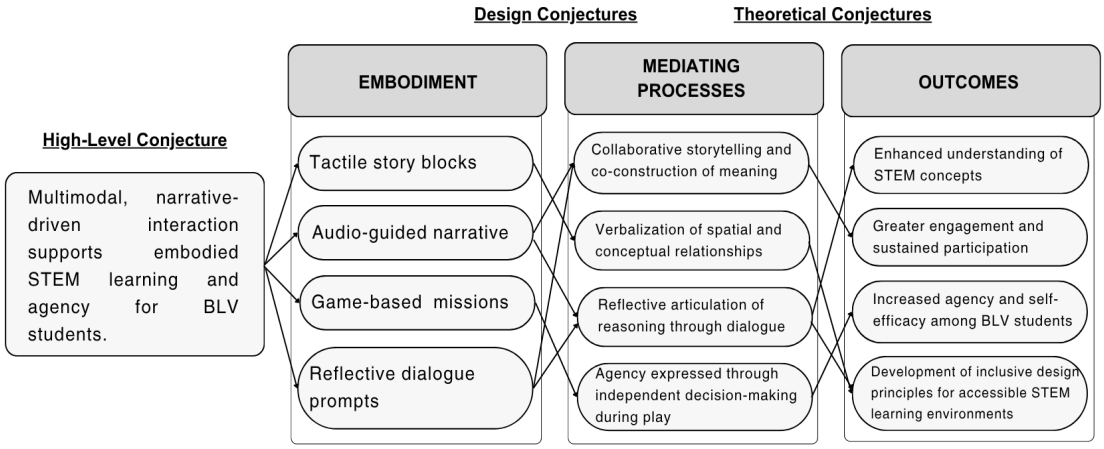
In addition to design, teachers stressed a shared responsibility for inclusion. They argued that general education teachers should integrate BLV students into lessons, rather than relying solely on TVIs to adapt instruction. As Ms. Cruz (25 years of experience) observed: *"Too often, classroom teachers assume that TVIs will handle everything, rather than making their own instruction more inclusive from the beginning. I would love for teachers to provide better verbal descriptions of what they're writing or displaying on screen."*

These findings suggest that inclusive design must extend beyond technological innovation to include shared pedagogical responsibility and classroom practices that foster learner independence, ensuring that accessibility translates into equitable participation. Technologies such as DotQuest aim to reinforce learners' *competence* through scaffolded challenges, adaptive feedback, so students' confidence grows when technology supports mastery of the concepts rather than dependency.

Discussion and conclusions

This study advances understanding of how multimodal, game-based environments can foster equitable STEM participation for BLV learners. The DotQuet design provides accessible practice opportunities while fostering motivation and engagement through progress-based, reward-driven structures similar to those commonly experienced by sighted students in digital learning environments (Liu et al., 2023). By situating accessible design not only as a question of accommodation but as a driver of meaningful engagement, this study contributes to ongoing discussions around equity and inclusion in technology-enhanced STEM education (Pires et al., 2019; Uttamchandani, 2018). Through interviews with TVIs, we refined our design conjectures for DotQuest, identifying how synchronized tactile–audio feedback supports non-visual sensemaking, narrative quests sustain motivation and math literacy, and inclusive orchestration promotes learner independence (see Figure 1). Rather than a finished product, DotQuest functions as a technology probe that bridges theory and practice in accessible design. These findings lay the groundwork for the next DBR cycle, which will involve co-design and classroom implementation to examine how multimodal interaction can foster accessibility, agency, and well-being in authentic learning contexts.

Figure 1
The Conjecture Map for DotQuest for BLV Students.



References

Anderson, T., & Shattuck, J. (2012). Design-Based Research: A Decade of Progress in Education Research? *Educational Researcher*, *41*(1), 16-25. https://doi.org/10.3102/0013189X11428813

Ávila-Soto, M., Valderrama-Bahamóndez, E., & Schmidt, A. (2017). TanMath: A tangible math application to support children with visual impairment to learn basic arithmetic. *Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments (PETRA '17)* (pp. 244–245). Association for Computing Machinery. https://doi.org/10.1145/3056540.3064964

Chen, S., Cox, E., Koh, K., & Seo, J. (2024). The Role of Verbal Descriptions in Facilitating Computational Thinking Skills for BVI Learners. In Lindgren, R., Asino, T. I., Kyza, E. A., Looi, C. K., Keifert, D. T., & Suárez, E. (Eds.), *Proceedings of the 18th International Conference of the Learning Sciences* - ICLS 2024 (pp. 2343-2344). https://doi.org/10.22318/icls2024.594746

Foley, E. A., Liu, Y., & Wilkerson, M. H. (2025). Teachers of Students With Visual Impairments' Role in Supporting Nonvisual Spatial and Graphical Literacies for Blind and Low Vision Students. In Rajala, A., Cortez, A., Hofmann, R., Jornet, A., Lotz-Sisitka, H., & Markauskaite, L. (Eds.), *Proceedings of the 19th International Conference of the Learning Sciences* - ICLS 2025 (pp. 709-717). https://doi.org/10.22318/icls2025.663874

Hutchins, N. M., & Biswas, G. (2024). Co-designing teacher support technology for problem-based learning in middle school science. *British Journal of Educational Technology*, 55(3), 802-822. https://doi.org/10.1111/bjet.133633

Kim, J., & Yeh, T. (2015). Toward 3D-printed movable tactile pictures for children with visual impairments. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)* (pp. 2815–2824). https://doi.org/10.1145/2702123.2702144

Lang, F., Pues, V., Schmidt, A., & Machulla, T.-K. (2023). BrailleBuddy: A tangible user interface to support children with visual impairment in learning Braille. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)* (Article 218, pp. 1–14). https://doi.org/10.1145/3544548.3580844

Pires, A. C., Marichal, S., Gonzalez-Perilli, F., Bakala, E., Fleischer, B., Sansone, G., & Guerreiro, T. (2019). A tangible math game for visually impaired children. In *Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '19)* (pp. 670–672). Association for Computing Machinery. https://doi.org/10.1145/3308561.335459

Plass, J. L., Homer, B. D., & Kinzer, C. K. (2015). Foundations of Game-Based Learning. *Educational Psychologist*, *50*(4), 258–283. https://doi.org/10.1080/00461520.2015.1122533

Rose, D. (2000). Universal Design for Learning. *Journal of Special Education Technology*, *15*(4), 47-51. https://doi.org/10.1177/016264340001500407 (Original work published 2000)

Sandoval, W. (2013). Conjecture Mapping: An Approach to Systematic Educational Design Research. *Journal of the Learning Sciences, 23*(1), 18–36. https://doi.org/10.1080/10508406.2013.778204

Uttamchandani, S. (2018). Equity in the learning sciences: Recent themes and pathways. *International Society of the Learning Sciences*, Inc.[ISLS]. https://doi.dx.org/10.22318/cscl2018.480

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (Vol. 86). Harvard university press.

Wu, H., Yang, H., Chang, F., Zhu, D., & Liu, Z. (2025). AI-generated tactile graphics for visually impaired children: A usability study of a multimodal educational product. *International Journal of Human-Computer Studies*, 103525. https://doi.org/10.1016/j.ijhcs.2025.103525

Acknowledgments

This work was supported by a collaborative grant from the National Science Foundation (NSF) [#2445678] awarded to Maryam Etezadbrojerdi and Franceli L. Cibrian. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. We also acknowledge Brandon Foley and Jeffrey Kraskouskas for their ongoing contributions to the broader research efforts of the lab.

International Society of the Learning Sciences

Pieces of the Puzzle: Examining Multimodal Interactions in Two Jigsaw Tasks to Inform Collaboration Support

Indrani Dey, University of Wisconsin-Madison
Collette Heskett, University of Colorado Boulder
Mon-Lin Monica Ko, University of Colorado Boulder

Abstract: This study examined students' verbal and nonverbal collaboration across two jigsaw tasks with differing cognitive complexity. Findings show that combining modalities provides a more complete picture of group dynamics and that interaction patterns vary by task complexity. These insights inform the design of AI partners that adapt support to both evolving collaboration and the task's cognitive demands.

Background and motivation

Collaborative knowledge building depends on dialogic interactions (Mercer, 2000), building shared understanding (Roschelle & Teasley, 1995), and regulation (Järvelä & Hadwin, 2013). Sustaining these social and cognitive demands is challenging (Jarvenoja et al., 2019), creating a need for structured support. One widely used approach in CSCL is the jigsaw method (Aronson et al., 1978), which requires students to integrate distributed expertise into shared knowledge. While jigsaw can promote collaboration (Roseth et al., 2019), it can increase task complexity, yielding mixed outcomes (Vives et al., 2024), prompting the need for continued and adaptive support.

Recent advancements in AI have enabled adaptive systems that support collaborative learning by monitoring learners' cognitive and interactional states. Prior work shows that AI-based scaffolding can support students' learning (Winkler et al., 2020), engagement (Liu et al., 2025), reasoning (Adamson et al., 2014), and regulation (Järvelä et al., 2023). However, many systems provide limited insight into their underlying mechanisms, focus on individuals, or rely primarily on verbal data. However, effective collaboration requires coordination of verbal and nonverbal behaviors (Barron, 2000, 2003; Roschelle & Teasley, 1995), such as joint attention, which predicts collaborative success (Schneider & Pea, 2013). An overreliance on verbal indicators risks overlooking quieter but actively engaged students (Inigaki et al., 1991; Obenland et al., 2012), highlighting the need for AI systems that attend and respond to multimodal interaction in small-group learning.

Our study addresses this need by examining students' verbal and nonverbal behaviors across two jigsaw activities with the goal of highlighting how these behaviors might be leveraged by AI systems to support collaboration. Building on prior analyses of small-group interaction (Dey & Puntambekar, 2023) and visualization techniques (Dey et al., 2025), we extend this work to a larger set of groups and compare two lessons that differ in cognitive complexity. We address the following research questions:

RQ1. What are the differences in verbal and nonverbal indicators between two jigsaw tasks of different cognitive complexities?

RQ2. What are the patterns and relationships in verbal and nonverbal indicators?

RQ3. *What do we learn when verbal and nonverbal indicators are examined together?*

Methods

Participants and context

This study draws on data from the NSF National AI Institute for Student-AI Teaming, a multi-institutional effort to develop AI partners to support collaborative learning in secondary classrooms. In this study, we use data from an 8th-grade ELA classroom in the Western US that implemented an eight-lesson unit on human and AI moderation of online communities. Our analysis focuses on two jigsaw lessons. In Lesson 3 (L3), students examined different community agreements and determined their use in familiar online contexts, emphasizing divergent thinking to develop new ideas. In Lesson 4 (L4), students compared five moderation approaches, and collaboratively summarized and ranked them based on co-developed criteria to reach consensus on which best support inclusive and safe gaming communities. Since consensus-oriented tasks demand greater deliberation and consideration of multiple perspectives (Xia et al., 2021), L4 was considered to be more cognitively demanding.

Small groups of 2-4 students were recording using an iPad on a tripod attached to a Yeti microphone. The recordings were transcribed, separated by utterances, and chunked into 5-minute segments. We randomly selected and analyzed segments of four different groups as they engaged in group discussions in L3 and L4.

 © ISLS

Analytic process

We analyzed the data using both verbal (using transcripts) and nonverbal (using videos) coding schemes.

Verbal analysis

We developed the verbal codes (Table 1) using deductive and inductive approaches, and refined them by connecting them to the learning objectives of the unit. We examined a 10-second segment of the transcript (including multiple utterances) and assigned a binary code of present (1) or absent (0) to each code (not mutually exclusive. Two authors double-coded 20% of the transcripts for each group, achieving an 86.4% agreement. Discrepancies were resolved through discussions, and one author coded the remaining segments.

Codes captured task orientation (*on-task*/*off-task*), idea exchange (*sharing idea*s and *ignoring ideas*), negotiation, and social dynamics. *Negotiation* was differentiated as 'knowledge building' or 'metacognitive', reflecting either building shared understanding (Roschelle & Teasley, 1995) or coordinating roles and tasks (Järvelä & Hadwin, 2013). Consensus was coded as positive (*Consensus+*) when groups reached agreement or re-entered negotiation, or negative (*Consensus–*) when disagreement was left unresolved. Positive (*Respect+*) and negative (*Respect–*) respect reflected active acknowledgment or dismissal of peers (Sleebos et al., 2006), while psychological safety captured behaviors that encouraged (*building*) or discouraged (*reducing)* open idea sharing and participation (Edmondson, 1999; Miyake & Kirschner, 2014).

Table 1

Verbal Coding Scheme, with the Codes, their Working definitions, and Examples

Indicators	Codes	Working definition	Example
On-task	*ontask*	Group discussion is related to the activity.	*..AI is trained to detect keywords.*
Off-task	*offtask*	Group discussion is not related to the activity.	*..Olivia Rodrigo is top tier.*
Negotiation	*nego_M* *nego_KB*	Involving more than one person; two types: ● *Metacognitive*: Related to group tasks or roles ● *Knowledge Building*: Related to what knowledge is being negotiated.	*Hold on, what's it called? What's our approach?*
Consensus+ (positive)	*condos*	Reaching a unanimous decision (either agree or disagree)	*Okay*
Consensus- (negative)	*conneg*	Does not reach a unanimous decision. May express disagreement but move on without further discussion.	*No, I can tell you this one's the worst.*
Sharing ideas	*share_idea*	Thoughts relating to knowledge building	*I think..*
Ignoring ideas	*ignore*	Ideas proposed by individuals do not get addressed or responded to by any other group member.	*What was blue again?*
Respect+ (positive)	*respos*	Acknowledging group members, active listening, and providing affirming responses	*Good job. I love that.*
Respect- (negative)	*resneg*	Ignoring others, responding rudely, or otherwise engaging in behavior that fosters negative feelings within the group.	*I did not ask you for your full depth opinion.*
Building psychological safety	*psbuild*	Encouraging others to share ideas, actively listening to other's ideas, affirming other's ideas, respectfully stating positions on others' ideas, even if they disagree.	*[Name redacted] you have to say something.*
Reducing psychological safety	*psbreak*	Shaming, punishing, or rejecting group members for sharing ideas or opinions, or being disrespectful during disagreements.	*Okay, that's not very good grammar.*

Nonverbal analysis

We coded the nonverbal interactions using the NICE framework (Dey & Puntambekar, 2023). The NICE framework captures students' nonverbal behaviors when working in small groups along three modalities: *eye gaze* (joint or individual gaze), *hand motions* (use of tool and pointing), and *body movements* (lean toward and away, and head nods). We coded each student's nonverbal behaviors for each utterance, with a "1" indicating the presence of a nonverbal behavior and "0" indicating its absence. The codes were not mutually exclusive. Two authors double-coded at least 20% of the utterances for each video, achieving a 74.2% agreement. Discrepancies were resolved by reviewing videos and discussion; one author coded the remainder.

To provide a uniform metric that could be then used to compare occurrences of an indicator in a certain interval of time (approximately 10 seconds), we normalized the verbal and nonverbal codes by dividing the occurrences of the codes for a 10-second segment, by the number of utterances in that segment.

Correlation analysis

We generated boxplots to examine the distribution of verbal and nonverbal behaviors over both lessons. We averaged the nonverbal behaviors by group to generate the distribution of nonverbal behaviors by lesson. We conducted Pearson's correlations for the verbal codes and nonverbal codes, and generated a heatmap of the correlations of the verbal behaviors and nonverbal behaviors, individually. To examine verbal and nonverbal behaviors together, we aggregated:

- **Positive verbal (*verbal_pos*)** = Negotiation (metacognitive) + Negotiation (knowledge building) + Consensus+ (positive) + Sharing new ideas + Respect+ (positive) + Building psychological safety
- **Negative verbal (*verbal_neg*)** = Consensus- (negative) + Ignore ideas + Respect- (negative) + Breaking psychological safety
- **Joint behaviors (*nv_joint*)** = Joint visual attention + Look at others + Use common tool + Point at tool + Point at others + Head nods/shakes + Lean towards
- **Individual behaviors (*nv_ind*)** = Individual attention + Look away + Use individual tool + Lean away

We performed Pearson's correlations between these four codes to examine verbal and nonverbal together.

Results and discussion

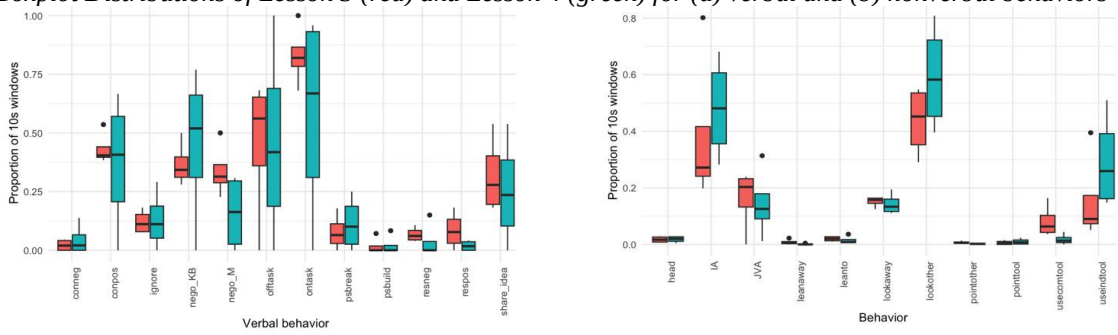
Verbal and nonverbal interaction patterns across L3 and L4

Verbal behaviors were more variable in Lesson 4 than Lesson 3 (Figure 1a). Students stayed on task more consistently in L3, though its open-ended structure also led to more off-task talk. L3 elicited more sharing of ideas, expressions of respect, and metacognitive negotiation, likely due to students' familiarity with the content. L4, which required integrating less familiar material, prompted more knowledge-building negotiation, indicating a need for support that helps groups regulate attention in more cognitively demanding tasks.

Individual behaviors such as individual attention (*IA*) and use of individual tools (*useindtool*) occurred more frequently in L4 (Figure 1b), while joint behaviors, such as joint visual attention (*JVA*), use of the common tool (*usecomtool*), and leaning toward others (*leanto*), were more frequent in L3. Although both lessons followed a jigsaw structure, the higher cognitive demands of L4 may have encouraged students to work independently.

Figure 1

Boxplot Distributions of Lesson 3 (red) and Lesson 4 (green) for (a) verbal and (b) nonverbal behaviors



(a) (b)

Correlations: Verbal and nonverbal indicators, individually

Figure 2(a) shows a heatmap of the verbal code correlations. On-task behavior was positively associated with productive collaboration, including knowledge-building and metacognitive negotiation, sharing ideas, and positive consensus, while off-task behavior showed negative associations with these processes, but was weakly linked to negative consensus, indicating that off-task talk rarely co-occurred with collaborative processes. Psychological safety showed weak relationships with task focus but *building* was positively associated with metacognitive negotiation and positive consensus. This suggests that feeling safe supports collaborative interactions, even if it occurs alongside off-task interactions. Negative respect was strongly linked to reduced psychological safety, indicating that disrespectful comments, rather than the absence of politeness, undermine trust and represent a key area for intervention. Knowledge-building negotiation was associated with both positive and negative consensus, suggesting that unresolved conflicts often center on conceptual issues, while idea sharing correlated with both deeper negotiation and ignored ideas, highlighting missed opportunities to build on student contributions. These patterns suggest that task design can shape not only task focus but also the quality of

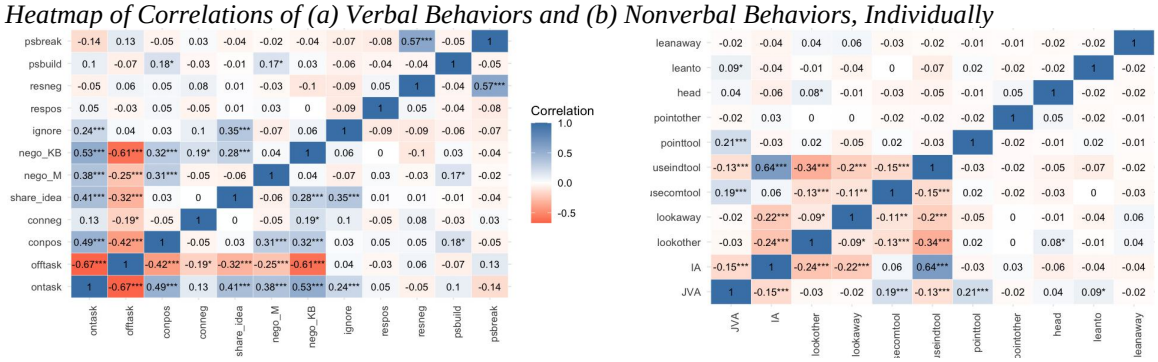
 © ISLS

collaboration, highlighting the need for supports that sustain attention, foster psychological safety, and help groups acknowledge and resolve conceptual disagreements.

Figure 2(b) presents the correlations among nonverbal behaviors. Joint visual attention (JVA) was positively associated with common tool use and indicators of joint engagement, such as pointing to the tool and leaning toward the activity area or peers, while individual attention was strongly linked to using individual tools. This pattern suggests that where students direct their attention is closely tied to how they engage with the task, highlighting the importance of support that promotes shared tool use to sustain mutual attention. JVA was not associated with looking at others or away, suggesting that shared attention can be maintained without frequent shifts in gaze. Head nods were weakly associated with looking at peers, indicating their role as nonverbal signals of attention and understanding. These findings highlight the value of coordinating gaze, gesture, and tool use to support shared engagement, particularly in tasks that risk students working more individually as is done in jigsaws.

Figure 2

Heatmap of Correlations of (a) Verbal Behaviors and (b) Nonverbal Behaviors, Individually



(a) (b)

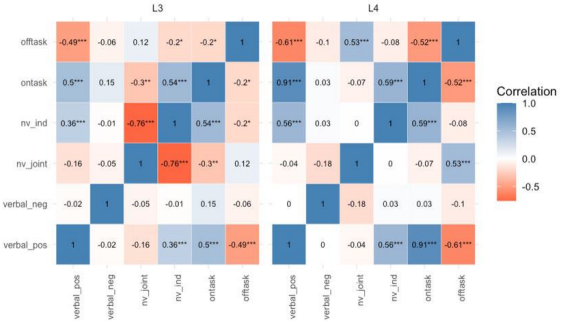
Note: p-value <=0.001:***, <=0.01: **, and <=0.05: *

Correlations: Verbal and nonverbal indicators, together

We examined the correlation between positive verbal behaviors (*verbal_pos*), negative verbal behaviors (*verbal_neg*), joint nonverbal behaviors (*nv_joint*), individual nonverbal behaviors (*nv_ind*), and on-task (*ontask*) and off-task (*offtask*) behaviors (Figure 3). Across both lessons, positive verbal behaviors were strongly associated with individual nonverbal engagement and on-task behavior, with stronger relationships in the more cognitively demanding L4. This pattern suggests that during harder tasks, students stayed focused through positive verbal exchanges but worked more independently rather than collaboratively. Positive verbal behaviors were not meaningfully linked to joint nonverbal coordination, indicating that students could remain engaged without aligning gaze or actions. Joint nonverbal behaviors were negatively associated with on-task behavior in L3 and positively associated with off-task behavior in L4, suggesting that shared gestures and mutual attention often occurred during off-task moments that may support social rapport rather than task progress. Individual nonverbal behaviors aligned with on-task focus, suggesting students tended to work individually despite the jigsaw structure.

Figure 3

Heatmap of Verbal and Nonverbal Indicators, Together



Note: p-value <= 0.001:***, <= 0.01: **, and <= 0.05:*.

Overall, our findings indicate that even in collaborative task designs (like jigsaw), additional support is needed to promote coordinated interaction and shared attention. Integrating verbal and nonverbal analyses reveals how task design might shape collaboration and highlights the importance of utilizing multimodal indicators to develop collaborative support.

 © ISLS

References

Adamson, D., Dyke, G., Jang, H., & Rosé, C. P. (2014). Towards an agile approach to adapting dynamic collaboration support to student needs. *International Journal of Artificial Intelligence in Education, 24*(1), 92-124.

Aronson, E., Stephan, C., Sikes, J., Blaney, N., & Snapp, M. (1978). The Jigsaw Classroom. *Sage Publication*.

Barron, B. (2000). Achieving coordination in collaborative problem-solving groups. *Journal of the Learning Sciences, 9*(4), 403–436. https://doi.org/10.1207/S15327809JLS0904_2

Barron, B. (2003). When smart groups fail. *Journal of the Learning Sciences, 12*(3), 307–359. https://doi.org/10.1207/S15327809JLS1203_1

Dey, I. & Puntambekar, S. (2023). Examining nonverbal interactions to better understand collaborative learning. In *Damșa, C., Borge, M., Koh, E., & Worsley, M. (Eds.), Proceedings of the 16th International Conference on Computer-Supported Collaborative Learning - CSCL 2023* (pp. 273-276). International Society of the Learning Sciences.

Dey, I., & Puntambekar, S. (2024). *The NICE framework: Examining nonverbal interactions in collaborative groups.* [Manuscript submitted]. Discourse Processes.

Dey, I., Ko, M. L. M., & Puntambekar, S. (2025). Visualizing Collaboration Using Multiple Modalities. In *Oshima, J., Chen, B., Vogel, F., & Järvelä, J. (Eds.), Proceedings of the 18th International Conference on Computer-Supported Collaborative Learning - CSCL 2025* (pp. 664-666). International Society of the Learning Sciences.

Edmondson, A. (1999). Psychological Safety and Learning Behavior in Work Teams. *Administrative Science Quarterly, 44*(2), 350–383. https://doi.org/10.2307/2666999

Järvelä, S., & Hadwin, A. F. (2013). New frontiers Regulating learning in CSCL. *Educational Psychologist, 48*(1), 25–39. https://doi.org/10.1080/00461520.2012.748006

Järvelä, S. , Nguyen, A. , & Hadwin, A. (2023). Human and artificial intelligence collaboration for socially shared regulation in learning. *British Journal of Educational Technology, 54*, 1057–1076. https://doi.org/10.1111/bjet.13325

Liu, Q., Yang, X., & Li, X. (2025, May). Promoting Student Engagement in CSCL Through Scaffolds and Generative AI-Based Conversational Agent. In *2025 5th International Conference on Artificial Intelligence and Education (ICAIE)* (pp. 113-116). IEEE.

Mercer, N. (2000). Words and minds: How we use language to think together. *Routledge*.

Miyake, N., & Kirschner, P. (2014). The social and interactive dimensions of collaborative learning. *The Cambridge Handbook of the Learning Sciences*, 418–438.

Obenland, C. A., Munson, A. H., & Hutchinson, J. S. (2012). Silent Students' Participation in a Large Active Learning Science Classroom. *Journal of College Science Teaching, 42*(2).

Roschelle, J., & Teasley, S. (1995). The construction of shared knowledge in collaborative problem solving. *Computer Supported Collaborative Learning*, 69-97. https://doi.org/10.1007/978-3-642-85098-1_5

Roseth, C. J., Lee, Y.-k., & Saltarelli, W. A. (2019). Reconsidering jigsaw social psychology: Longitudinal effects on social interdependence, sociocognitive conflict regulation, motivation, and achievement. *Journal of Educational Psychology, 111*(1), 149–169. https://doi.org/10.1037/edu0000257

Schneider, B., & Pea, R. (2013) Real-time mutual gaze perception enhances collaborative learning and collaboration quality. *International Journal of Computer-Supported Collaborative Learning, 8*(4), 375-397. https://doi.org/10.1007/s11412-013-9181-4

Sleebos, E., Ellemers, N., & de Gilder, D. (2006). The carrot and the stick: Affective commitment and acceptance anxiety as motives for discretionary group efforts by respected and disrespected group members. *Personality and Social Psychology Bulletin, 32*(2), 244-255.

Winkler, R., Hobert, S., Salovaara, A., Söllner, M., & Leimeister, J. M. (2020, April). Sara, the lecturer: Improving learning in online education with a scaffolding-based conversational agent. In *Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1-14).

Xia, T., Kang, M., Chen, M., Ouyang, J., & Hu, F. (2021). Design training and creativity: Students develop stronger divergent but not convergent thinking. *Frontiers in Psychology, 12*, 695002.

Acknowledgments

We thank the teacher and student participants, and our team of data collectors, transcribers, and annotators for their support. This research has been supported by the National Science Foundation (Grant # 2019805).

International Society of the Learning Sciences

Reconceptualizing Activity Theory for Human-AI Teaming in Computer-Supported Collaborative Learning

Jaeyoon Choi, University of California, Irvine, jaeyoon.choi@uci.edu
Nia Nixon, University of California, Irvine, dowelln@uci.edu

Abstract: The rise of artificial intelligence (AI) and large language models (LLMs) are reshaping the role of technology in Computer-Supported Collaborative Learning (CSCL). Once viewed as mediating tools, AI systems can now act as active teammates that co-construct meaning, share responsibility, and influence team cognition. This paper addresses the lack of theoretical grounding in existing human-AI teaming research by reconceptualizing Engeström's Activity Theory. Specifically, we extend the six components of the activity system (subject, tool, rules, community, division of labor, and object) to account for AI's dual role as both tool and quasi-subject within collaborative learning activities. Through this theoretical mapping, we highlight emerging questions for future research, including but not limited to issues of learner agency, trust, equity, and epistemic alignment. We conclude with a call for the CSCL community to revisit and expand its theoretical frameworks to better capture the evolving dynamics of human-AI teaming.

Introduction

While technology has long been a central component of Computer-Supported Collaborative Learning (CSCL), it has traditionally functioned as a *mediating artifact*, facilitating communication, coordination, and shared understanding among human learners without directly contributing to the discourse (Jeong & Hmelo-Silver, 2012). However, recent advancement in artificial intelligence (AI), particularly in large language models (LLMs), have fundamentally transformed this dynamic. That is, AI is no longer confined to the role of a *passive* tool but can now participate as an *active* teammate, co-constructing meaning, generating ideas, and contributing to both cognitive and social dimensions of collaboration. This emerging paradigm, also known as human-AI teaming, reconceptualizes the very foundations of CSCL, as AI teammates begin to assume team-level responsibilities, engage in decision-making processes, and influence the direction and quality of collaborative learning.

Despite the growing body of research on human-AI teaming, much of the current work has been technologically or empirically driven, focusing on design innovations (e.g., Chen et al., 2025), task performance (e.g., Flathmann et al., 2023), and user engagement and trust (e.g., Duan et al., 2025). For example, Choi et al. (2025) examined how AI teammates can reshape socio-cognitive dynamics of collaborative problem solving by examining discourse collected from experiments. While human-AI teaming research remains a nascent domain within CSCL, these studies often lack a comprehensive theoretical framework for understanding how collaboration with AI teammates transform the dynamics of group interaction, knowledge co-construction, and goal orientation. Without such conceptual grounding, we risk treating AI merely as an additional actor within existing collaborative systems rather than interrogating how its presence may reorganize the social, cognitive, and cultural structures of learning. In other words, there is a pressing need for a framework that captures the systemic, mediated, and evolving nature of Human-AI teaming dynamics.

To address this gap, this paper reconceptualizes a seminal theory used to understand collaboration in CSCL, namely Activity Theory. Originated from the socio-cultural work of Vygotsky and Leontiev and later expanded by Engeström, Activity Theory provides a systemic lens for analyzing how human actions are mediated by cultural tools, social rules, and collective goals (Barab et al., 2013). In CSCL research, Activity Theory has been used to examine how interactions among learners, teachers, and mediating artifacts shape the process of collaboration and knowledge construction (Collis & Margaryan, 2004; Hmelo-Silver & Chernobilsky, 2012). Building on this, we extend Activity Theory to conceptualize AI not merely as a tool but as a *quasi-subject* that participates in and transforms the collaborative activity system. Specifically, we aim to map the six components of Engeström's activity system – subject, tools, rules, community, division of labor, and object – to the context of human-AI teaming, and to compare how these relationships and tensions within the activity system differ from those in traditional human-human collaboration. Through this reconceptualization, we aim to highlight how the integration of AI teammates in CSCL reorganizes the fundamental structure and dynamics of collaborative learning activity.

The subsequent sections of the paper are organized as follows. We first provide a brief overview of human-AI teaming and Activity Theory, focusing on Engeström's expansion of the framework. We then present our conceptual mapping of Activity System components to human-AI teaming, highlighting key differences from

 © ISLS

traditional human-human collaboration. Finally, we discuss emerging research directions and critical questions for future inquiry into human-AI collaboration in learning contexts.

Human-AI teaming

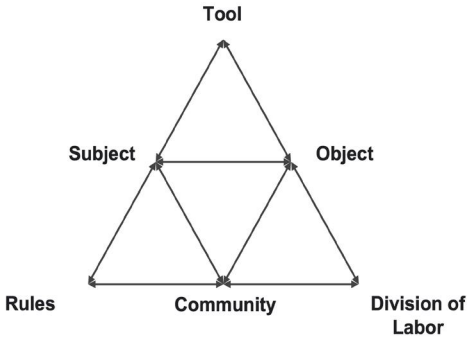
The rapid advancement of AI and LLMs has changed how humans interact with computers. In traditional CSCL, computers functioned as mediating artifacts that facilitated communication and coordination among learners, but now, AI is no longer confined to the role of a tool; it can act as an active teammate. This emerging paradigm, often referred to as human-AI teaming, emphasizes that AI agents can assume team-level responsibilities, engage in decision-making, and contribute meaningfully to shared goals (Harris-Watson et al., 2023; Seeber et al., 2020). While still nascent, early empirical work in human-AI teaming shows that integrating AI teammates can enhance both team performance and learning outcomes (Bienefeld et al., 2024). For instance, Bienefeld et al. (2024) found that physicians and nurses working with AI in a simulated clinical setting demonstrated stronger hypothesis generation and speaking-up behaviors. Moreover, Schelble et al. (2025) showed that human-AI teams can develop team cognition over time, influenced by team members' positive attitudes toward their AI teammates. Despite these promising empirical findings, much of human-AI teaming work remains primarily empirical or design-oriented, with limited theoretical insight into how the presence of AI reshapes social and cognitive dimensions of collaboration.

Activity Theory

Activity Theory originated within the socio-cultural tradition of psychology, offering a framework for understanding the human mind as inherently social, object-oriented, and mediated (Engeström, 1999). Rooted in the work of Vygotsky and Leontiev, Activity Theory challenges the individualistic view of learning by situating cognition within collective activity systems. Specifically, Vygotsky emphasized that human development occurs through mediated action, where individuals interact with their environment not directly but through cultural and historical artifacts such as language, tools, and symbols (Vygotsky, 1978). Leontiev further expanded this idea by conceptualizing activity as the essential unit of analysis to understand psychological processes (Leontiev, 1981). Building on this foundation, Engeström developed a systemic model of activity that contextualized mediated action within collective processes. According to Engeström, activity is the minimal meaningful unit of analysis for understanding human behaviors – that is, it is through activity that humans develop skills, consciousness, transform social conditions, resolve contradictions, and generate new cultural tools. His model is often depicted as a triangular representation, comprising six interrelated components: subject, tool, object, rules, community, and division of labor (see Figure 1). These components together form a dynamic and interdependent system where change in any component affects the other. More importantly, contradictions or tensions that arise within or between components are viewed as drivers of innovation and learning, facilitating the evolution of both the activity and its participants.

Figure 1

Engeström's Activity Theory model.



In CSCL, Activity Theory has been widely applied to analyze how learners, teachers, and mediating artifacts co-construct knowledge and shape collaborative processes. For instance, Hmelo-Silver & Chernobilsky (2012) used Activity Theory to examine students' learning processes in an online collaborative environment, where groups jointly analyze video cases and applied learning sciences concepts to lesson design. Similarly, Collis & Margaryan (2004) demonstrated how Activity Theory can be used to understand CSCL in workplace-oriented

 © ISLS

contexts. However, the application of Activity Theory in CSCL has primarily focused on human-human collaboration, where technology serves as a passive mediating tool rather than an active participant.

Reconceptualization of Activity Theory in human-AI teaming

In this section, we map the six components of Activity Theory onto the context of human-AI teaming, highlighting how each element differs from traditional human-human collaboration within CSCL settings. Our goal is to identify emerging tensions, possibilities, and critical research questions that arise when AI systems participate as teammates in collaborative learning environments. Note that these questions are not meant to be exhaustive; rather, they aim to foreground key areas of questions that are particularly relevant to the CSCL community and the broader field of education.

Subject

In traditional human-human collaboration, the subject typically denotes an individual learner or a group of learners participating in a shared activity. In human-AI teaming context, this notion of subject still remains focused on the human learners, because our primary interest is in how human learners engage, think, and learn in tandem with AI and other human teammates. However, the presence of an AI teammate introduces a new challenge – learners now oscillate between treating the AI as a tool and as a peer collaborator, which gives rise to novel tensions and affordances. One central question is learner agency: do learners feel they retain control over the interaction, or do they perceive themselves as subordinate to the AI's suggestions or decisions? For instance, Darvishi et al. (2024) found that students using AI assistance during peer feedback tasks tended to rely on the AI's response rather than learning from it, suggesting risks of reduced agency. Similarly, cognitive offloading to AI may lower cognitive burden but also lead to dependency; for instance, Gerlich (2025) observed a negative correlation between frequent AI tool use and critical thinking, mediated by cognitive offloading, thereby raising questions about how shared cognition with AI teammates might shape learning processes and outcomes. Finally, trust is a crucial factor influencing the degree to which learners engage with AI partners – whether human learners perceive an AI teammate as competent, reliable, and aligned with shared goals determines the depth of collaboration and the quality of outcomes. Georganta & Ulfert (2024) emphasize that perceived trustworthiness and similarity of AI teammates play a central role in fostering effective team trust.

Tool

In traditional CSCL, tools, especially computers and online collaborative platforms, are conceived as mediating artifacts that facilitate communication, coordination, and knowledge construction. In human-AI teams, however, the AI teammate now occupies a dual role: it continues to act as a mediating artifact (e.g., by providing suggestions, generating content and visual representations, or how students increasingly treat AI as a search engine replacement (Divekar et al., 2025)) while simultaneously functioning as a *quasi-subject* that participates in meaning-making and decision-making. This dual role challenges the conventional subject-tool boundary and prompts key design questions. First, should the AI teammate be visibly distinct from human, or should we make it disguise as human (Seeber et al., 2020)? Research on anthropomorphism suggests that adding human-like design cues can increase perceived trustworthiness (Ma et al., 2025), yet anthropomorphizing can blur judgement about the AI system in general and introduce ethical concerns (Placani, 2024). These tensions also extend to the question of AI persona – what kind of persona and/or identity, if any, should an AI teammate display?

Rules

In both traditional CSCL and human-AI teams, rules encompass the formal and informal norms that govern the activity within a community. These could include classroom/course expectations, institutional policies, and group norms/dynamics that shape collaboration, such as turn-taking conventions or feedback etiquette. While the concept of rules remains consistent in human-AI teaming contexts, their manifestations change when an AI teammate joins the team. First, it introduces a new ethical question: what moral or ethical code should guide the AI's behavior? For instance, recent efforts such as Anthropic's Constitutional AI framework (Bai et al., 2022) highlights attempts to embed normative principles directly into AI systems. At the team level, designers and instructors should also consider what social norms the AI teammate should follow. For example, an AI might be prompted to follow equal turn-taking to promote equity in participation. However, surface-level equality in participation might not necessarily equate to meaningful inclusion, and this could also hinder authentic discourse (Nixon et al., 2024). Finally, as AI systems increasingly generate ideas and shape collective decisions, the question of accountability becomes critical: who is responsible for the outputs and actions that emerge from human-AI teaming – the human teammates, the instructor, or the system designer?

Community

The community in Activity Theory refers to the group of individuals whose knowledge, interests, stakes, and goals shape the activity. In both traditional CSCL and human-AI teaming contexts, the community typically consists of the collaborating team members engaged in a shared task. A central concern within these communities is group cohesion – the sense of solidarity and bond linking members of a team together. In human-AI teams, however, it remains unclear whether the presence of an AI teammate strengthens or undermines this cohesion. For instance, Flathmann et al. (2023) found that competing motivations between human and AI teammates can disrupt collaborative processes, suggesting that AI teammates' differing goals may impact group cohesion. These dynamics invite further questions about sense of belonging and inclusion: does an AI teammate foster meaningful connection, or might it inadvertently diminish human members' sense of participation and emotional engagement?

Division of labor

Division of labor refers to how tasks and responsibilities are distributed within the team. Beyond just who does which task, we can also explore socio-cognitive role distribution: the social roles (e.g., facilitating interaction, maintaining group climate) versus cognitive roles (e.g., executing task-related reasoning, computations) (Graesser et al., 2018). In human-AI teams, it is likely that an AI teammate will excel in cognitive roles, given its computational speed and content-generation capacity. For instance, studies such as Choi et al. (2025) found that AI teammates tend to assume leadership in cognitive facilitation (e.g., guiding the discussion, suggesting next steps). However, this dominance in cognitive activity can result in overreliance on AI, where humans might become passive information consumers rather than active collaborators. Lastly, social roles pose a critical design challenge: can AI meaningfully contribute to social cohesion and emotional alignment in ways that foster collaboration? Future CSCL research should explore how AI can support rather than suppress social engagement.

Object

The object represents the shared problem space or goal that directs collaborative activity. In both traditional CSCL and human-AI teaming contexts, it corresponds to the collaborative task that learners work on together. However, when an AI teammate is involved, the nature and design of the task itself warrant careful consideration. One of the questions we should ask is what kinds of learning activities benefit from the integration of human and AI knowledge and capabilities (Seeber et al., 2020). Moreover, evaluating progress in human-AI teams requires moving beyond traditional measures of task completion or correctness. Instead, assessment should also consider the quality of collaboration, the distribution of cognitive contributions, and the degree of epistemic alignment between human and AI teammates – in other words, how well the team co-constructs understanding throughout the activity (Cress & Kimmerle, 2023).

Conclusion: Call for CSCL community

In this paper, we reconceptualized Activity Theory, a foundational socio-cultural framework in CSCL, to understand how the activity system can be reimagined in the era of human-AI teaming. By mapping the six core elements of Engeström's activity system – subject, tool, roles, community, division of labor, and object – we explored how the introduction of AI teammates reconfigures established relationships within collaborative learning. This theoretical mapping revealed new forms of tension, opportunity, and design considerations: from learner agency and trust, to the dual role of AI as a tool and a quasi-subject, to emerging questions about equity, sense of belonging, and accountability in human-AI teams.

While Activity Theory provides a comprehensive framework for understanding human-AI teaming dynamics, it is not the only one capable of doing so. Theories such as distributed cognition can also offer important insights into the social, cognitive, and technological dimensions of human-AI teaming. Hence, we encourage the CSCL community to revisit and expand its theoretical foundations to better understand how humans and AI learn and work together. With the rapid advancement of AI in collaborative learning, we now encounter new interactional phenomena, data, and methods that invite us to examine which aspects of existing theories remain explanatory and which begin to break down when AI becomes a teammate.

Moving forward, we urge CSCL researchers to conceptualize AI not merely as a mediating tool but as an active teammate – one that can shape meaning, regulate interaction, and influence the socio-cognitive and affective dynamics of collaboration. Until we recognize and theorize AI in this way, our analyses will continue to underestimate its impact on learning processes and outcomes. We call on the CSCL community to take a leading role in this theoretical expansion of human-AI teaming. The integration of AI into collaborative learning challenges us to reconceptualize the field's foundational principles, and in turn, build the next generation of theory-informed, ethically grounded, and human-centered approaches to collaborative learning in the age of AI.

References

Bai, Y., Kadavath, S., Kundu, S., Askell, A., Kernion, J., Jones, A., Chen, A., Goldie, A., Mirhoseini, A., McKinnon, C., & others. (2022). Constitutional ai: Harmlessness from ai feedback. *arXiv Preprint arXiv:2212.08073*.

Barab, S. A., Evans, M. A., & Baek, E.-O. (2013). Activity theory as a lens for characterizing the participatory unit. In *Handbook of research on educational communications and technology* (pp. 208–223). Routledge.

Bienefeld, N., Keller, E., & Grote, G. (2024). Human-AI teaming in critical care: A comparative analysis of data scientists' and clinicians' perspectives on AI augmentation and automation. *Journal of Medical Internet Research*, *26*, e50130.

Chen, Z., Luo, Y., & Sra, M. (2025). Engaging with AI: How Interface Design Shapes Human-AI Collaboration in High-Stakes Decision-Making. *arXiv Preprint arXiv:2501.16627*.

Choi, J., Samadi, M. A., JaQuay, S., Park, S., & Nixon, N. (2025). *Read the Room or Lead the Room: Understanding Socio-Cognitive Dynamics in Human-AI Teaming*. https://arxiv.org/abs/2510.09944

Collis, B., & Margaryan, A. (2004). Applying activity theory to computer-supported collaborative learning and work-based activities in corporate settings. *Educational Technology Research and Development*, *52*(4), 38–52.

Cress, U., & Kimmerle, J. (2023). Co-constructing knowledge with generative AI tools: Reflections from a CSCL perspective. *International Journal of Computer-Supported Collaborative Learning*, *18*(4), 607–614.

Darvishi, A., Khosravi, H., Sadiq, S., Gašević, D., & Siemens, G. (2024). Impact of AI assistance on student agency. *Computers & Education*, *210*, 104967.

Divekar, R. R., Gonzalez, L., Guerra, S., Boos, N., & Divekar, R. (2025). Can Generative AI Replace Search Engines for Learning? Understanding Student Preferences, Use and Perceived Proficiency in Using AI. *TechTrends*, 1–16.

Duan, W., Flathmann, C., McNeese, N., Scalia, M. J., Zhang, R., Gorman, J., Freeman, G., Zhou, S., Hauptman, A. I., & Yin, X. (2025). Trusting Autonomous Teammates in Human-AI Teams-A Literature Review. *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, 1–23.

Engeström, Y. (1999). Activity theory and individual. *Perspectives on Activity Theory*, 19.

Flathmann, C., Schelble, B. G., Rosopa, P. J., McNeese, N. J., Mallick, R., & Madathil, K. C. (2023). Examining the impact of varying levels of AI teammate influence on human-AI teams. *International Journal of Human-Computer Studies*, *177*, 103061.

Georganta, E., & Ulfert, A.-S. (2024). Would you trust an AI team member? Team trust in human–AI teams. *Journal of Occupational and Organizational Psychology*, *97*(3), 1212–1241.

Gerlich, M. (2025). AI tools in society: Impacts on cognitive offloading and the future of critical thinking. *Societies*, *15*(1), 6.

Graesser, A. C., Fiore, S. M., Greiff, S., Andrews-Todd, J., Foltz, P. W., & Hesse, F. W. (2018). Advancing the science of collaborative problem solving. *Psychological Science in the Public Interest*, *19*(2), 59–92.

Harris-Watson, A. M., Larson, L. E., Lauharatanahirun, N., DeChurch, L. A., & Contractor, N. S. (2023). Social perception in Human-AI teams: Warmth and competence predict receptivity to AI teammates. *Computers in Human Behavior*, *145*, 107765.

Hmelo-Silver, C. E., & Chernobilsky, E. (2012). Understanding Collaborative Activity Systems: The Relation of Tools and Discourse in Mediating Learning. In *Embracing Diversity in the Learning Sciences* (pp. 254–261). Routledge.

Jeong, H., & Hmelo-Silver, C. (2012). *Technology supports in CSCL*.

Leontiev, A. N. (1981). The problem of activity in psychology. In J. V. Wertsch (Ed.), *The concept of activity in Soviet psychology* (pp. 120–142). M. E. Sharpe.

Ma, N., Khynevych, R., Hao, Y., & Wang, Y. (2025). Effect of anthropomorphism and perceived intelligence in chatbot avatars of visual design on user experience: Accounting for perceived empathy and trust. *Frontiers in Computer Science*, *7*, 1531976.

Nixon, N., Lin, Y., & Snow, L. (2024). Catalyzing equity in STEM teams: Harnessing generative AI for inclusion and diversity. *Policy Insights from the Behavioral and Brain Sciences*, *11*(1), 85–92.

Placani, A. (2024). Anthropomorphism in AI: hype and fallacy. *AI and Ethics*, *4*(3), 691–698.

Schelble, B. G., Flathmann, C., Macdonald, J. P., Knijnenburg, B., Brady, C., & McNeese, N. J. (2025). Modeling perceived information needs in human-AI teams: Improving AI teammate utility and driving team cognition. *Behaviour & Information Technology*, *44*(9), 2069–2092.

Seeber, I., Bittner, E., Briggs, R. O., De Vreede, T., De Vreede, G.-J., Elkins, A., Maier, R., Merz, A. B., Oeste-Reiß, S., Randrup, N., & others. (2020). Machines as teammates: A research agenda on AI in team collaboration. *Information & Management*, *57*(2), 103174.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard university press.

Acknowledgments

This research was supported in part by the Jacobs Foundation Grant #2024-1533-00, Spencer Vision Grant # 202600129 and the Bill & Melinda Gates Foundation Grant # INV-068818.

International Society of the Learning Sciences

Designing Digital Enhancements for Social Emotional Skill in a Collaborative Learning Environment

Enruo Guo, Emily Kerzabi, Jessica Andrews-Todd, Jiangang Hao, Patrick Kyllonen
eguo@ets.org, ekerzabi@ets.org, jandrewstodd@ets.org, jhao@ets.org, pkyllonen@ets.org
ETS Research Institute

Stephen Elliott, Arizona State University, steve@ssiscolab.com

Abstract: This study presents the design and implementation of a fifth-grade Social and Emotional Learning (SEL) curriculum within a hybrid digital collaborative environment. Using a user-centered design approach, we engaged teachers and a school administrator as co-designers over a four-month period to iteratively refine the features and tools from a collaborative learning and assessment system (EPCAL). Key design decisions, such as tailoring class analytics to SEL and integrating feedback and rewards systems, were informed by teacher input. A semester-long pilot study involving two teachers and 22 students demonstrated the feasibility of delivering SEL through a hybrid model using EPCAL. Findings highlight strong teacher engagement and meaningful student participation. This research contributes a scalable framework for hybrid SEL delivery, demonstrating how a participatory design approach can be effectively applied to curriculum and technology enhancements - an approach that can be generalized to other hybrid learning contexts.

Introduction

Social and Emotional Learning (SEL) is a foundational component of education, equipping students with essential skills such as emotional regulation, empathy, relationship-building, and responsible decision-making. These competencies support healthy identity development and help students navigate complex social environments (CASEL, 2021). However, teaching and assessing SEL remains challenging due to its reliance on interpersonal interaction and collaboration - elements that traditional classroom methods often struggle to capture effectively.

Hybrid or blended learning environments, where students engage both in-person and through digital platforms, offer new opportunities to support SEL instruction. This format allows for flexible lesson delivery and classroom management, while enabling students to connect through multiple modalities. These settings are ideal for SEL's social and experiential nature, while also enabling data collection to benefit teachers, students, and researchers. To support SEL in this context, we adapted the Social Skills Improvement System Classwide Intervention Program (SSIS-CIP; Elliot & Gresham, 2020), grounded in the CASEL framework (Borowski, 2019; Jagers et al., 2019). This program outlines five core competencies: self-awareness, self-management, social awareness, relationship skills, and responsible decision-making. SSIS-CIP is traditionally delivered through a teacher-led slide deck with scripted activities, where teachers cannot get the detailed interactions of students in each team. To address this limitation, we introduced a computer-supported collaborative learning (CSCL) solution designed to enhance both instructional delivery and assessment, where students still participate in physical activities like role plays and discussions, but also engage in small-group chats via an on-screen interface (Hao et al., in press). This CSCL solution is facilitated by the ETS Platform for Collaborative Assessment and Learning (EPCAL; Hao et al., 2017) comprises two core components: a learner portal that facilitates collaborative tasks, displays performance data, and integrates teacher feedback; and a teacher portal that enables team management, real-time monitoring of student interactions, and post-lesson review of behavioral and performance indicators. Adapting EPCAL for SEL instruction in hybrid classrooms presented four key design challenges:

- First, the teacher portal, originally developed for remote collaboration, required greater flexibility to accommodate in-person classroom dynamics, such as late arrivals, early departures, and weak internet connectivity issues.
- Second, the existing analytics dashboard lacked SEL-specific metrics and feedback mechanisms, limiting its utility for formative assessment.
- Third, feedback from prior studies indicated a need for a more engaging and user-friendly interface to better support student and teacher interaction.
- Finally, the curriculum itself required adaptation to fit the hybrid format, including the integration of group discussions, self-reflection prompts, and generalization questions into the digital system, along with revised teacher scripts to guide their implementation.

 © ISLS

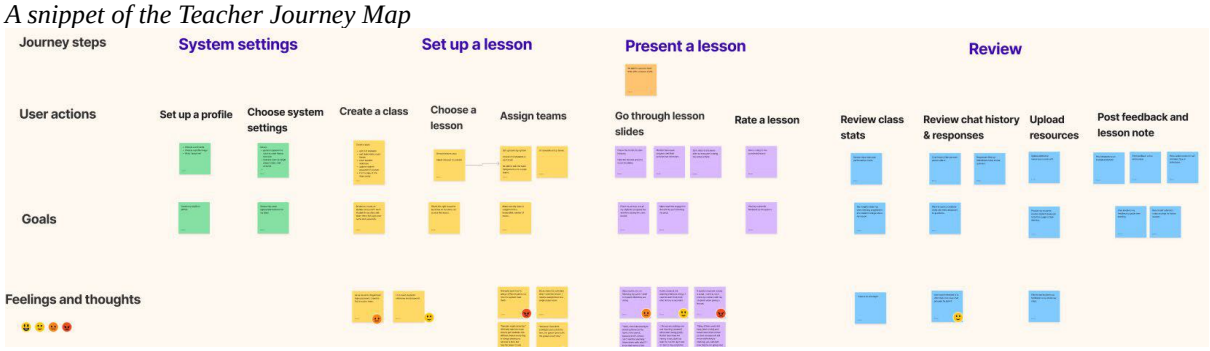
Design process

A Design-Thinking approach

Designing for hybrid classrooms requires a multi-faceted strategy that integrated human-centered design and UI/UX principles to better understand user needs and improve system usability. We adopted the five-stage design thinking framework - *Empathize*, *Define*, *Ideate*, *Prototype* and *Test* (Hasso Plattner Institute of Design at Stanford University, n.d.), to guide our process. During the *Empathize* and *Define* stages, we conducted contextual research and cognitive lab studies with teachers and students in a summer program, gathering initial feedback on the existing system. These insights informed the development of user personas, journey maps, and prioritized feature sets. In the *Ideate* phase, we generated user stories, explored classroom scenarios, and defined user flows to address key challenges. The *Prototype* and *Test* phases involved iterative co-design sessions with teachers, allowing us to refine the system collaboratively. A particularly impactful method was mapping the teacher journey within EPCAL (see Figure 1). Using a collaborative visualization board to capture key steps, emotions, and pain points significantly improved design efficiency and fostered a shared understanding of the features needed to support collaborative learning.

Figure 1

A snippet of the Teacher Journey Map



The teacher journey map highlighted the main steps, goals and actions in delivering a SEL lesson, as well as teacher's feelings and thoughts, pain points and opportunities to mitigate them. Before teaching a lesson, the teacher configures system settings, selects a profile image, creates a class, adds student accounts, chooses a lesson, and assigns students to teams of 2–4. During the lesson, they use the presentation tool to guide slides and manage classroom dynamics, such as removing offline students, restarting teams, or freezing students' screens. Teachers can monitor real-time student conversations and review class performance afterward. Feedback can be given to the whole class or individual students.

Co-Design

During *Prototype* and *Test* phase, we adopted a co-design approach (Dollinger et al., 2019; Dollinger & Lodge, 2018; Sanders & Stappers, 2008) by inviting two teachers from our summer cognitive lab study and a district-level school administrator to join our research team as co-designers. This collaboration began in Fall 2024 and continued through monthly sessions until the end of the year. Under the co-design model, researchers and designers work closely with end users to iteratively shape the final product. Each session included one hour of synchronous discussion and approximately 30 minutes of asynchronous tasks for participants, such as surveys or preparatory work for the next meeting. Developers, UI/UX designers, and content experts contributed additional time between sessions to refine prototypes and incorporate feedback. Teachers were encouraged to envision ideal features and workflows, which were then prototyped, tested, and reimagined collaboratively. This iterative process not only accelerated development and reduced redundancy but also deepened our understanding of how the system would be used in practice. Below we highlight two main design decisions derived from co-design with teachers.

Class analytics dashboard

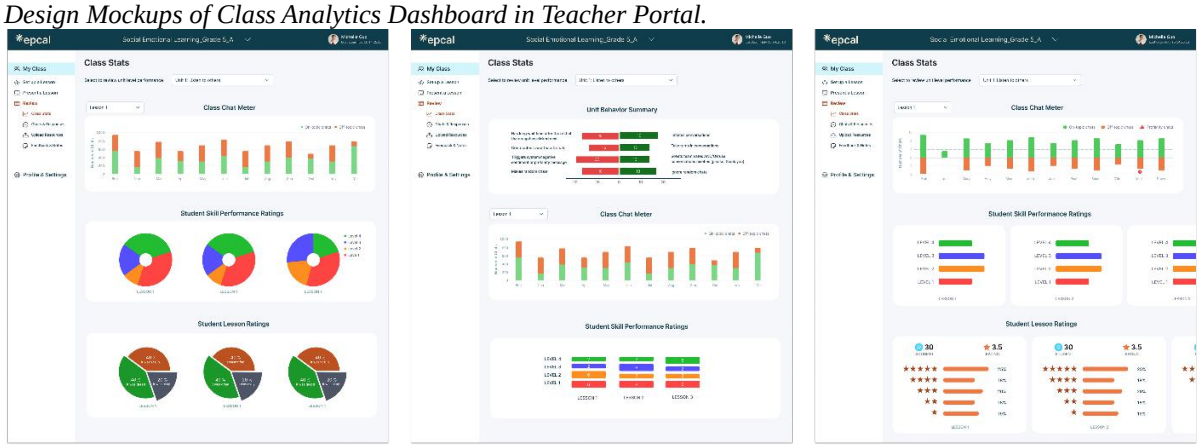
One of the primary challenges within the teacher portal was the need to tailor the class analytics dashboard to the SEL domain. To address this, we presented teachers with low-fidelity wireframes and mock-up charts, prompting them to reflect on the types of insights they would expect from such a dashboard. Teachers emphasized the importance of conversation-based analytics, such as the frequency of specific keywords, indicators of productive versus non-productive chat, and individual contributions to group activities. They also noted that certain metrics, like participation rates or lesson timing, offered limited value in assessing SEL development. Three key

 © ISLS

components: class chat analysis, student skill performance ratings, and lesson ratings, were derived through co-design. The presentation format and level of detail for each component underwent multiple design iterations. (Figure 2).

Figure 2

Design Mockups of Class Analytics Dashboard in Teacher Portal.



From left to right, this figure illustrates the evolution of the class analytics page within the teacher portal. All three designs feature a class chat meter - a vertical bar graph showing on-topic (green) and off-topic (orange) chats. The rightmost design adds detection of profanity, while the middle design includes deeper analytics like sentiment, turn-taking, and conversation initiation, which some teachers find overwhelming. Student skill ratings have evolved from a donut chart (left) to horizontal bar charts (middle and right), better illustrating level progression. Lesson ratings have shifted from a pie chart to a star-bar chart, aligning with familiar rating conventions.

Feedback and rewards system

In a separate co-design session, we focused on the teacher feedback loop and the integration of a rewards system. Using a Think-Aloud protocol, teachers walked through the process of completing a lesson, reviewing student responses and chat interactions. They also went through the feedback page where they can provide class-level or individual feedback. To support student engagement, we proposed a set of digital stickers as age-appropriate rewards. Teachers provided input on the relevance and visual design of the stickers, as well as their placement within the system. We discussed two integration options (see Figure 3). (1) Awarding stickers during real-time chat interactions, and (2) embedding stickers within the feedback workflow. Teachers favored the latter, noting that it aligned more intuitively with their existing feedback practices and avoided overwhelming the chat interface.

Pilot study

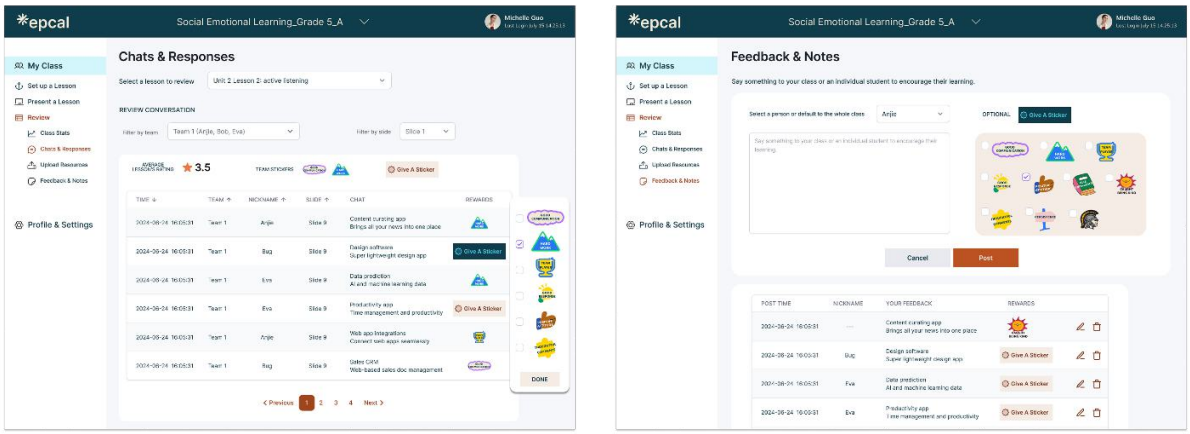
In Spring 2025, we conducted a four-month pilot study in Northern Michigan (U.S.) involving two teachers and 22 students in 5th grade. Across this period, teachers delivered 36 lessons covering 12 SEL skill units (e.g., “Listening to others,” “Taking turns when you talk”). Each lesson followed six instructional events: Tell, Show, Do, Practice, Monitor, and Generalize. Students worked in small groups (2-4 members) to engage in collaborative SEL activities such as group chats, role plays, reflection, and generalization exercises. All student interactions, responses, and activities were recorded through EPCAL, enabling teachers to monitor engagement in real time and review behavioral and performance data after the lesson. We partnered with a local school administrator to observe classroom implementation and collect teacher and student feedback through observation notes, while also analyzing system log data to track usage patterns. The pilot study focused on five key dimensions of technology-enhanced curriculum delivery: (1) classroom setup and technology integration, (2) classroom management and teacher engagement, (3) student engagement, (4) role play implementation, and (5) lesson disruptions.

Findings indicated that class sizes ranged from 6 to 16 students, with flexible layouts, such as pods, rows, and circles, to support interaction and role play. Teachers consistently demonstrated strong engagement, using the system to manage classroom dynamics including late arrivals and early departures. Larger classes experienced more frequent technical issues, such as logouts and chat failures, though overall system reliability improved over time. Despite these challenges, students were generally thoughtful and engaged, showing increasing comfort in expressing ideas and participating in discussions. Role play remained a central instructional strategy, though its effectiveness varied depending on clarity of instructions and transition management. Disruptions, primarily due to student movement and technical issues, were handled efficiently by teachers to maintain lesson flow.

During the pilot, we sent surveys to teachers about curriculum implementation and use of the sticker feature. Feedback on curriculum modifications highlighted concerns about the chat feature, which some teachers found disruptive to pacing. Students occasionally misused the chat, e.g., excessive emojis or switching keyboard languages, leading to delays and peer frustration. While students generally appreciated the chat function, its misuse prompted targeted design changes to improve system messaging, streamline teacher tools, and promote appropriate communication within the student portal. Both teachers found the sticker feature easy to use and effective for providing individual feedback. However, they expressed differing views on its usefulness when issuing stickers to the entire class. Although they used the feature occasionally, they mentioned that it effectively supported student engagement and reinforced positive behaviors.

Figure 3

Design Mockups for Integrating Stickers into the Teacher Portal.



Left: Award stickers while reviewing student group chats, embedding rewards in real-time collaboration. Right: Integrate stickers into the feedback workflow, allowing teachers to award them alongside written comments to individuals or the class.

Conclusions

Overall, the design practices implemented in this project, specifically co-design and teacher partnerships, significantly advanced the development of a hybrid classroom model that integrates curriculum and technology enhancements to support SEL. Looking ahead, incorporating role plays and other interactive modalities into the digital platform holds promise for creating more immersive learning experiences and enabling more authentic assessments of student performance. These efforts lay the groundwork for scalable, data-informed SEL instruction that is both engaging and pedagogically sound.

References

Borowski, T. (2019). *CASEL's framework for systemic social and emotional learning*. Establishing Practical Social-Emotional Competence Assessment Work Group. https://measuringsel.casel.org/wp-content/uploads/2019/08/AWG-Framework-Series-B.2.pdf

Collaborative for Academic, Social, and Emotional Learning (CASEL) (2021). *Fundamentals of SEL*. https://casel.org/fundamentals-of-sel/

Dollinger, M., Liu, D., Arthars, N., & Lodge, J. M. (2019). Working together in learning analytics towards the co-creation of value. *Journal of Learning Analytics, 6*(2), 10–26. https://doi.org/10.18608/jla.2019.62.2

Dollinger, M., & Lodge, J. M. (2018). Co-creation strategies for learning analytics. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge* (pp. 97–101). https://doi.org/10.1145/3170358.3170372

Elliott, S.N., & Gresham, F.M. (2020). SSIS SEL Classwide Intervention Program: Training, Implementation, & Evaluation. SAIL CoLab.

Hao, J., Liu, L., von Davier, A. A., Lederer, N., Zapata-Rivera, D., Jakl, P., & Bakkenson, M. (2017). EPCAL: ETS platform for collaborative assessment and learning. *ETS Research Report Series, 2017*(1), 1–14.

Hao, J., Kyllonen, P., Kerzabi, E., Cui, W., Guo, E., Andrews-Todd, J., Jiang, Y., & Elliott, S. (in press). Leveraging CSCL for social emotional learning: Early insights from a large-scale classroom study. Paper accepted to the International Conference on Computer-Supported Collaborative Learning (CSCL 2026).

Hasso Plattner Institute of Design at Stanford. (n.d.). *Design Thinking Bootleg.* https://dschool.stanford.edu/tools/design-thinking-bootleg

Jagers, R. J., Rivas-Drake, D., & Williams, B. (2019). Transformative social and emotional learning (SEL): Toward SEL in service of educational equity and excellence. *Educational Psychologist, 54*(3), 162–184. https://doi.org/10.1080/00461520.2019.1623032

Sanders, E. B.-N., & Stappers, P. J. (2008). Co-creation and the new landscapes of design. *CoDesign, 4*(1), 5–18. https://doi.org/10.1080/15710880701875068

Acknowledgments

This work was funded in part by the Education Innovation Research (EIR) program of the Office of Elementary and Secondary Education (OESE), U.S. Department of Education, (#S411C230179). The views, opinions, and/or findings contained in this paper are those of the authors and shall not be construed as an official position, policy, or decision of the funding agencies, unless so designated by other documents.

International Society of the Learning Sciences

Linking Physiological Data and Self-Reports: Exploring Engagement in Collaborative Learning

Xiaoran Li, Beijing Language and Culture University, lxr77@blcu.edu.cn
Serkan Ucan, Istanbul Medeniyet University, serkan.ucan@medeniyet.edu.tr

Abstract: This study investigated the relationship between physiological synchrony and engagement in an authentic collaborative learning context. Heart rate data and self-reported engagement surveys were collected from 32 postgraduate students during a semester-long collaborative writing course. Data were analyzed using Monte Carlo shuffling, Spearman correlation, and Welch's t-test. Spearman correlation revealed no significant associations between heart rate-based physiological synchrony and any dimension of self-reported engagement. However, pairs exhibiting significant physiological synchrony reported notably higher socio-emotional engagement compared to non-synchronous pairs, with this difference approaching statistical significance. These findings suggest that in authentic learning environments characterized by dynamic complexity, a direct linear relationship between physiological synchrony and engagement may not always be evident. Nevertheless, when physiological synchrony reaches a statistically significant threshold, it appears to serve as a sensitive indicator of the socio-emotional dimensions of collaboration, highlighting its value for uncovering complex dynamics of student engagement in naturalistic educational settings.

Introduction

Collaborative learning has been recognized as an important pedagogical approach due to its potential to foster learners' collaborative problem-solving, critical thinking, and cooperative competence (Li et al., 2023, 2025). However, effective collaboration does not automatically emerge from merely grouping learners together; it requires their sustained engagement with the task and one another (Haataja et al., 2025; Skinner & Pitzer, 2012). Student engagement is generally conceptualized as commitment or effortful involvement in learning (Henrie et al., 2015). It is a multidimensional construct encompassing behavioral, cognitive, and socio-emotional dimensions (Fredricks et al., 2004; Xu et al., 2023). Behavioral engagement involves observable effort and participation; cognitive engagement reflects deep learning strategies; and socio-emotional engagement captures learners' affective states and the quality of their social interactions (Fredricks et al., 2004). Due to its importance for learning, researchers have continuously sought to investigate engagement in collaborative learning to reveal its characteristics.

Traditionally, researchers use post-hoc self-report questionnaires or observations to study engagement in collaborative learning (Appleton et al., 2006; Henrie et al., 2015). While valuable, these methods have limitations: self-reports provide only static accounts, and observations are labor-intensive, difficult to scale, and cannot fully capture learners' internal cognitive and affective states (Haataja et al., 2025). These limitations have led to calls for more objective, instantaneous, and efficient measures of engagement (Zhang et al., 2021). With advances in biosensing technologies, physiological measures such as skin conductance and heart rate have been introduced to explore engagement (Haataja et al., 2025; Hammond et al., 2023; Stuldreher et al., 2023). However, this line of research remains early and findings are inconsistent, with most studies conducted in laboratory settings (Zhang et al., 2021). To address these gaps, this study examines the relationship between heart rate–based physiological synchrony and self-reported engagement in an authentic collaborative writing course. Despite the small sample size, it provides preliminary evidence and supports further research.

Literature review

Physiological synchrony in collaborative learning

Learning is an active process that leaves a physiological footprint—for example, elevated heart rate or increased skin-conductance responses frequency (Haataja et al., 2018). These indicators are key measures of the Autonomic Nervous System (ANS), whose Sympathetic (SNS) and Parasympathetic (PNS) branches work in tandem, dynamically changing as they regulate the body in preparation for and in response to environmental conditions (Palumbo et al., 2017). Extensive studies indicate that ANS responses are closely linked to learning-relevant processes, including enhanced emotion, increased attention, and better memory (Kreibig, 2010). As learning is becoming increasingly social, this motivates a shift in focus from intrapersonal ANS activity to interpersonal

 © ISLS

autonomic physiology (IAP), which is defined as the relationship between individuals' physiological dynamics, as indexed by continuous measures of the autonomic nervous system (Palumbo et al., 2017). Within this framework, physiological synchrony (PS) is a key indicator, referring to group members' similar, simultaneous changes in physiological signals, e.g., electrodermal activity (EDA) and heart rate (HR) (Dindar et al., 2020). Although ANS activity is a well-validated measurement of the learning process, it does not indicate valence or context. Therefore, secondary measures are needed to jointly interpret their meaning and to illuminate the learning process (Palumbo et al., 2017). Many studies continue in this vein, assuming that these interpersonal physiological processes can reflect invisible social processes co-occurring with observable interactions (Haataja et al., 2018). For example, Dindar et al. (2020) linked dyadic physiological synchrony derived from electrodermal activity with self-reported cognitive, motivational, and emotional self-regulation in learning, and found that physiological synchrony was only significantly related to cognitive self-regulation.

Physiological synchrony and engagement

Only recent studies have begun using physiological data to investigate engagement in collaborative learning, suggesting it offers a continuous and implicit measure compared to traditional approaches. Typically, such studies employ continuous signal analysis methods to compute dyadic or group-level physiological synchrony indices, which are then compared with engagement assessed through other measures to examine their relationships. For instance, Hammond et al. (2023) found that inter-subject heart rate (HR) synchrony was associated with self-reported attentional and emotional engagement, whereas no such relationship was observed for skin conductance. Conversely, Haataja et al. (2025) reported a negative association between electrodermal activity (EDA) synchrony and perceived value of collaboration, suggesting that high synchrony does not always indicate productive engagement.

However, research in this area remains in its early stages, and further studies are needed to validate these findings. Therefore, the present study was conducted in an authentic semester-long academic writing course to examine the relationship between group members' physiological synchrony and self-reported engagement. This study is guided by the following two research questions:

RQ1: How is group members' physiological synchrony associated with their self-reported behavioral, cognitive, socio-emotional, and overall engagement?

RQ2: How do the self-reported behavioral, cognitive, socio-emotional, and overall engagement levels differ between pairs with significant physiological synchrony and those without?

Method

Participants

This study was conducted at a normal university in northern China, involving 32 first-year postgraduate students majoring in education (mean age = 22.3 years, SD = 0.4; 6 male). All students were enrolled in an English academic writing course and participated voluntarily after being informed of the study's purpose. They were assured of confidentiality and the right to withdraw at any stage. The students self-organized into ten collaborative groups consisting of three to four members each.

Collaborative activity

The study was conducted in a semester-long, two-credit compulsory English academic writing course for postgraduate students majoring in education. In addition to weekly lectures, students were required to work in groups to produce a 4,000–6,000-word academic paper on a self-selected topic. Based on prior teaching and writing experience, the instructor encouraged each group to engage in at least one hour of face-to-face collaboration per week, accumulating a minimum of eight hours of joint work over the semester. Due to time constraints, groups were permitted to use previously collected but unpublished experimental data for the writing task. During the eight-week writing period, the groups were responsible for identifying a suitable dataset, formulating research questions, and collaboratively composing their paper. In each collaborative writing session, group members used their own computers to log in to Tencent Online Shared Document, where they jointly developed and revised their manuscript.

Data collection

Before the course began, participants' prior experience in reading and publishing English academic papers was examined. The results indicated that participants had comparable prior knowledge, with no significant differences among them.

 © ISLS

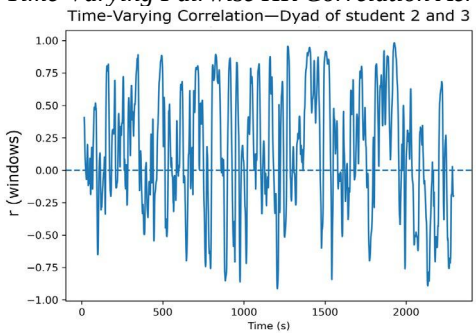
During each collaborative writing session, participants' heart rate and self-reported engagement data were collected. Each participant's heart rate (HR) was continuously and unobtrusively recorded using the Scosche Rhythm+ armband heart rate monitor (Valencell, Inc., Raleigh, NC, USA). The Scosche Rhythm + is a precision biometrics device designed to monitor HR. To collect data, the device was installed on each participant's non-dominant arm, and HR data were sampled at a frequency of 1 Hz. In addition, students' self-reported engagement was collected immediately after each session using the Collaborative Learning Engagement Questionnaire (Xu et al., 2023), a validated five-point Likert-scale instrument developed for the Chinese higher education context. The questionnaire measures three dimensions of engagement—behavioral, cognitive, and socio-emotional—and demonstrates high reliability, with Cronbach's α coefficients ranging from 0.82 to 0.98 for the subscales and 0.93 for the overall instrument.

Data analysis

Due to missing physiological and questionnaire data for some group members, and the absence of HR recordings in several sessions, this study obtained valid data for 19 of the 75 sessions. Preprocessing of the HR signal followed a commonly used workflow (e.g., see Dindar et al., 2020; Haataja et al., 2018). First, the HR data were visually inspected for clear signs of movement artifacts. Second, each participant's HR series was Z-score standardized to neutralize the individual differences among the students. Subsequently, following prior methods (e.g., Dindar et al., 2020; Haataja et al., 2018), physiological synchrony (PS) was estimated pairwise within each collaborating group by computing a Single Session Index (SSI) for each pair in every session. A sliding-window procedure (30-s window; 1-s step) was applied to obtain pairwise Pearson correlations between members' HR signals, yielding a time-resolved correlation series per pair for the session (Figure 1). SSI was computed by taking the natural logarithm of the ratio between (i) the sum of positive correlations and (ii) the sum of the absolute values of negative correlations across the session. Larger positive SSI values reflect stronger pairwise PS over the session. Accordingly, a three-person group yields three pairwise SSIs, and a four-person group yields six per session. Monte Carlo shuffling was further utilized to determine the significance of synchrony (e.g., see Dindar et al., 2020). Only actual SSI values higher than the highest shuffled concordance at $p < .05$ were considered to present significant synchronous pairs.

Figure 1

Time-Varying Pairwise HR Correlation Across the Session



To address RQ1, Spearman correlation was utilized to examine the relationship between SSI and the dyadic averages of self-reported cognitive, behavioral, socio-emotional, and overall engagement scores (Dindar et al., 2020). For RQ2, pairs were first classified as either synchronous (n=7) or non-synchronous (n=56) based on the Monte Carlo shuffling test. Then, the average self-reported engagement scores between these two groups were compared. Given the unequal group sizes, this study utilized Welch's t-test for the comparison, as it does not assume equal variances (Welch,1947). To further assess the robustness of the findings, this study also calculated and reported bias-corrected and accelerated (BCa) 95% bootstrap confidence intervals (Efron & Tibshirani, 1993).

Results

1. How is group members' physiological synchrony associated with their self-reported behavioral, cognitive, socio-emotional, and overall engagement.

Spearman correlation analysis showed no significant associations between physiological synchrony and any of the four self-reported engagement measures: behavioral engagement ($\rho = .103$, $p = .420$), cognitive engagement ($\rho = .056$, $p = .666$), socio-emotional engagement ($\rho = .110$, $p = .392$), and the overall score ($\rho = .120$, $p = .347$, $N = 63$).

2. How do the self-reported behavioral, cognitive, socio-emotional, and overall engagement levels differ between pairs with significant physiological synchrony and those without.

 © ISLS

Descriptively, synchronous pairs reported higher mean engagement across all dimensions (Table 1). While Welch's t-tests were not significant, the socio-emotional dimension approached significance (t=−1.95, p=.073). A more robust bootstrap analysis confirmed this difference was significant, as the BCa 95% confidence interval for the mean difference, [-3.51, -0.06], excluded zero.

Table 1
Differences in Engagement Scores Between Non-Synchronous and Synchronous Pairs

Dimension	Non-sync (n = 56) Mean ± SD	Sync (n = 7) Mean ± SD	Welch's t	p	Hedges' g	Bootstrap 95% CI for Mean Difference
Behavioral	44.63 ± 3.14	46.14 ± 3.53	-1.09	0.313	-0.47	[-4.25, 1.37]
Cognitive	45.34 ± 3.12	46.71 ± 2.22	-1.47	0.175	-0.45	[-3.34, 0.46]
Socio-emotional	46.09 ± 3.96	47.86 ± 1.95	-1.95	**0.073**	-0.46	[-3.51, -0.06]
Overall	136.05 ± 8.88	140.71 ± 7.16	-1.58	0.151	-0.53	[-10.75, 1.07]

Discussion

The present study links physiological data and self-reports in an authentic collaborative writing context to examine engagement from a physiological perspective, revealing a complex pattern that may extend existing research.

Overall, the findings revealed no significant association between heart rate–based physiological synchrony and any dimension of self-reported engagement. However, when significant synchrony occurred between pairs, their self-reported socio-emotional engagement was higher than in pairs without significant synchrony, with this difference approaching statistical significance. This finding diverges from some previous studies that reported positive relationships between physiological synchrony and some facets of engagement, such as cognitive engagement (Zhang et al., 2021), or both cognitive and socio-emotional engagement (Stuldreher et al., 2023). The results of this study suggest that, in authentic collaborative learning contexts, a direct linear relationship between physiological synchrony and engagement may not always be evident. Unlike lab-based research (e.g., Hammond et al., 2023; Stuldreher et al., 2023) —where task demands and social dynamics are more limited— real classrooms involve fluctuating engagement and diverse interaction styles (Haataja et al., 2025). Under such variability, physiological synchrony and engagement may not consistently manifest as simple linear relationships (Haataja et al., 2025). This study also uncovered that when physiological synchrony exceeded a statistical threshold (Monte-Carlo shuffling, $p < .05$) and achieved significance, dyads tended to report higher socio-emotional engagement. This finding reaffirms that significant physiological synchrony captures meaningful moments that warrant attention during collaborative processes (Dindar et al., 2020; Palumbo et al., 2017). Additionally, physiological synchrony appears more sensitive to socio-emotional aspects of collaboration than to cognitive or behavioral dimensions, consistent with prior research (e.g., Li et al., 2023). This may be because socio-emotional processes—encompassing learners' emotions and interaction quality (Fredricks et al., 2004)—are inherently interpersonal, and physiological synchrony, as an index of shared autonomic activity (Palumbo et al., 2017), is well suited to capturing these affective and relational dynamics.

Conclusion and limitation

This study examined the relationship between heart rate–based physiological synchrony and self-reported engagement in an authentic collaborative writing course. Although no significant linear correlations were found, dyads with significant physiological synchrony reported higher socio-emotional engagement than non-synchronous dyads, with the difference approaching significance. This suggests that physiological synchrony may reflect meaningful affective and relational aspects of collaboration. By linking physiological data with self-reports in a naturalistic setting, this study contributes to understanding engagement from a physiological perspective. Several limitations should be noted. The imbalance between synchronous (n = 7) and non-synchronous pairs (n = 56) limits statistical power. Although robust methods (e.g., Welch's t-test and bootstrapping) were used, this imbalance—common in authentic classroom contexts—requires cautious interpretation. Future research should validate these findings with larger, richer multimodal evidence across diverse contexts and examine potential non-linear relationships between physiological synchrony and engagement using advanced modeling techniques.

 © ISLS

References

Appleton, J. J., Christenson, S. L., Kim, D., & Reschly, A. L. (2006). Measuring cognitive and psychological engagement: Validation of the student engagement instrument. *Journal of School Psychology*,*44*(5), 427–445

Dindar, M., Malmberg, J., Järvelä, S., Haataja, E., & Kirschner, P. A. (2020). Matching self-reports with electrodermal activity data: Investigating temporal changes in self-regulated learning. *Education and Information Technologies*, *25*(3), 1785-1802.

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of educational research*, *74*(1), 59-109.

Henrie, C. R., Halverson, L. R., & Graham, C. R. (2015). Measuring student engagement in technology-mediated learning: A review. *Computers & Education*, *90*, 36-53.

Haataja, E., Malmberg, J., & Järvelä, S. (2018). Monitoring in collaborative learning: Co-occurrence of observed behavior and physiological synchrony explored. *Computers in Human Behavior*, *87*, 337-347.

Haataja, E., Törmänen, T., Somerville, M. P., Malmberg, J., Järvenoja, H., & Järvelä, S. (2025). A momentary view of engagement in collaborative learning: Triangulation through multimodal data. *Frontline Learning Research*, *13*(2), 102 – 121.

Hammond, H., Armstrong, M., Thomas, G. A., & Gilchrist, I. D. (2023). Audience immersion: validating attentional and physiological measures against self-report. *Cognitive Research: Principles and Implications*, *8*(1), 22.

Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological psychology*, *84*(3), 394-421.

Li, X., Hu, W., Li, Y., & Mao, Z. (2023). Exploring what synchronized physiological arousal can reveal about the social regulatory process in a collaborative argumentation activity. *Frontiers in psychology*, *13*, 1042970.

Li, X., Li, Y., Mao, Z., & Gong, R. (2025). The effects of group awareness support combined with collaboration scripts on adaptive social regulation in promoting students' challenge awareness and social regulation strategies. *Computers & Education*, *235*, 105324.

Palumbo, R. V., Marraccini, M. E., Weyandt, L. L., Wilder-Smith, O., McGee, H. A., Liu, S., & Goodwin, M. S. (2017). Interpersonal autonomic physiology: A systematic review of the literature. *Personality and social psychology review*, *21*(2), 99-141.

Stuldreher, I. V., van Erp, J. B., & Brouwer, A. M. (2023). Robustness of physiological synchrony in wearable electrodermal activity and heart rate as a measure of attentional engagement to movie clips. *Sensors*, *23*(6), 3006.

Skinner, E. A., & Pitzer, J. R. (2012). Developmental dynamics of student engagement, coping, and everyday resilience. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 21–44). Springer.

Tibshirani, R. J., & Efron, B. (1993). An introduction to the bootstrap. *Monographs on statistics and applied probability*, *57*(1), 1-436.

Welch, B. L. (1947). The generalization of ‘STUDENT'S’problem when several different population varlances are involved. *Biometrika*, *34*(1-2), 28-35.

Xu, B., Stephens, J. M., & Lee, K. (2024). Assessing student engagement in collaborative learning: development and validation of new measure in China. *The Asia-Pacific Education Researcher*, *33*(2), 395-405.

Zhang, J., Wang, K., & Zhang, Y. (2021). Physiological characterization of student engagement in the naturalistic classroom: A mixed-methods approach. *Mind, Brain, and Education*, *15*(4), 322-343.

International Society of the Learning Sciences

GenAI Chatbot for Socially Shared Regulation of Learning in CSCL: Impact on University Students' Learning Engagement and Outcome

Dengkang Chen, Yi Zhang, Jianwen Sun, Yijie Lu, Xiaoge Guo, Fucheng Guo
1300191113@qq.com, zhangyi@mail.ccnu.edu.cn, sunjw@ccnu.edu.cn, yijie_lu@mails.ccnu.edu.cn, guoxiaoge1025@qq.com, guofucheng@mails.ccnu.edu.cn
Faculty of Artificial Intelligence in Education, Central China Normal University
Laboratory for Artificial Intelligence and New Forms of Education, Central China Normal University

Abstract: Supporting socially shared regulation of learning (SSRL) is crucial for successful collaborative learning but is challenging due to its multidimensional and context-sensitive nature. While generative AI (GenAI) chatbots have shown promise in addressing these challenges, research has largely treated SSRL and GenAI chatbots separately, with limited empirical exploration of their integration. Furthermore, existing studies on GenAI chatbots supporting SSRL are mainly conceptual or design-focused, lacking empirical evaluation. Addressing these gaps, we developed SSRLBot, a GenAI chatbot designed to identify SSRL challenges and offer contextual strategies. We conducted a six-week experimental study with 55 Chinese university students, dividing them into two classes: one using SSRLBot (N = 27) and the other without chatbot support (N = 28). Learning engagement and outcomes were assessed through questionnaires, platform log data and products. Results showed that SSRLBot significantly improved students' learning engagement and outcomes, extending the application of GenAI chatbots from content tutor to group moderators in CSCL.

Introduction and literature review

Socially shared regulation of learning (SSRL) is a collective metacognitive process essential for successful collaboration, involving the co-construction of task understanding, shared goal setting, monitoring, adaptation, and reflection (Hadwin et al., 2011). Despite its importance, effectively scaffolding SSRL in Computer-supported collaborative learning (CSCL) remains challenging due to its dynamic and context-dependent nature (Kim et al., 2025). Traditional static scaffolds often lack the adaptability to meet groups' real-time regulatory needs, leading to fragmented or superficial regulation (Zheng et al., 2025; Zhou & Tsai, 2023).

While Generative AI (GenAI), particularly large language models (LLMs), presents a promising approach for adaptive and dialogic SSRL support, its potential is often underutilized. First, most efforts leverage GenAI for content-centric tutoring, largely overlooking its potential as a process-oriented team moderator (Dara et al., 2024). Second, even where regulation support is attempted, most studies remain conceptual or design-oriented, underscoring the need for more empirical research to establish their effectiveness. (Huang et al., 2025). Furthermore, these attempts are often fragmented, failing to scaffold the full SSRL process.

Addressing these gaps, the present study developed a GenAI chatbot for scaffolding SSRL (named SSRLBot) and conducted a six-week experimental study to explore its impact on university students' learning engagement and outcomes in CSCL. The following research questions (RQ) were formulated.

RQ1: What are the effects of selected GenAI chatbot on university students' learning engagement?

RQ2: What are the effects of GenAI chatbot on university students' collaborative learning outcomes?

Learning engagement in collaborative learning

Learning engagement refers to the quality of effort students invest in learning to achieve desired outcomes (Fredricks et al., 2004). In CSCL, engagement is shaped by social interaction and is commonly conceptualised as a multidimensional construct, which includes (Fredricks et al., 2004; Sinha et al., 2015): (a) Behavioral engagement, which refers to participation in collaborative tasks and involvement in academic activities. (b) Cognitive engagement, which refers to the knowledge construction process that occurs as group members work toward a common goal. Previous research has identified that the ICAP framework (Interactive, Constructive, Active, Passive) provides a valuable lens for analyzing cognitive engagement levels. According to the framework, interactive and constructive cognitive activities are most likely to result in higher levels of engagement, as they involve deeper processing and meaningful collaboration (Chi & Wylie, 2014). (c) Emotional engagement represents students' affective reactions about their collaborative experience and peers.

Despite the pivotal role of engagement, existing literature identifies several critical challenges in authentic CSCL settings, notably uneven behavioral involvement, shallow cognitive processing, and negative emotional experience (Chen et al., 2024). As a growing body of research links effective SSRL to high-quality

 © ISLS

engagement, scaffolding SSRL is increasingly recognized as a critical strategy to address these collaborative challenges (Isohätälä et al., 2017).

Supporting SSRL in collaborative learning with GenAI Chatbots

SSRL extends self-regulated learning (SRL) from the individual to the group level. Whereas SRL concerns learners' proactive regulation via cognitive, motivational, and behavioral strategies, SSRL emphasizes collective metacognitive regulation of shared cognition, motivation, and emotion toward common goals (Hadwin et al., 2011; Zimmerman, 1990). It is commonly conceptualized as a cyclical process of shared task understanding, shared goal setting and planning, shared monitoring and adapting, and shared reflecting (Panadero & Järvelä, 2015). Empirical studies have suggested that groups exhibiting high-level SSRL tend to achieve deeper engagement and produce higher-quality collaborative outcomes (Li et al., 2021).

However, effective SSRL rarely develops spontaneously in CSCL. Prior work has therefore examined scaffolds such as strategy instruction or question-based prompts across SSRL phases, which can improve engagement, outcomes, and reflective quality (Zheng et al., 2025; Zhou & Tsai, 2023). Despite these advances, limitations remain. Most scaffolds are insufficiently adaptive to groups' evolving needs and rarely address group heterogeneity or role differences. Additionally, they offer limited interactivity for co-constructing understanding and iteratively adjusting strategies.

Recent advances in large language models (LLMs) make GenAI chatbots a promising direction because they can sustain multi-turn, context-specific dialogue and offer bidirectional, just-in-time support aligned with core SSRL processes (Edwards et al., 2025). Building on this potential, emerging studies have begun designing SSRL-oriented GenAI chatbots. For example, Huang et al. (2025) introduced an LLM-based SSRLBot that summarizes group dialogue and generates theory-aligned evaluations and recommendations. Related CSCL interventions have embedded GenAI chatbots into collaborative platforms (e.g., PyramidApp) and reported improved answer accuracy and indications of better individual–group interaction quality (Sánchez-Reina et al., 2025). Nevertheless, several gaps remains. First, the evidence base is still limited, with many studies being conceptual or design-oriented rather than empirical. Moreover, few studies directly test whether SSRL-focused chatbots enhance learning engagement, and existing implementations often address only parts of the SSRL cycle. It is timely to develop a GenAI chatbot aligned with the full SSRL cycle and empirically examine its effects on engagement and learning outcomes in CSCL.

Methodology

Participants and procedure

The study was conducted in a six-week Educational Research Methods course at a Chinese university, involving 55 Educational Technology undergraduates (22 male, 33 female; aged 18–20) from two parallel classes with comparable prior knowledge. One class was randomly assigned to the experimental condition (SSRLBot-CSCL;), utilizing a GenAI chatbot to scaffold SSRL on the Feishu platform, while the other served as the control (T-CSCL;) without chatbot support. The experiment followed a three-stage design: (1) a pre-intervention week for platform orientation and pretesting; (2) a four-week intervention where both groups attended 90-minute lectures and collaborated on Feishu platform to develop group research proposals, with the experimental group receiving embedded SSRLBot scaffolding; and (3) a final week for proposal submission, presentations, and emotional engagement surveys. All participants provided informed consent following institutional ethical guidelines.

SSRLBot: a GenAI chatbot supporting SSRL

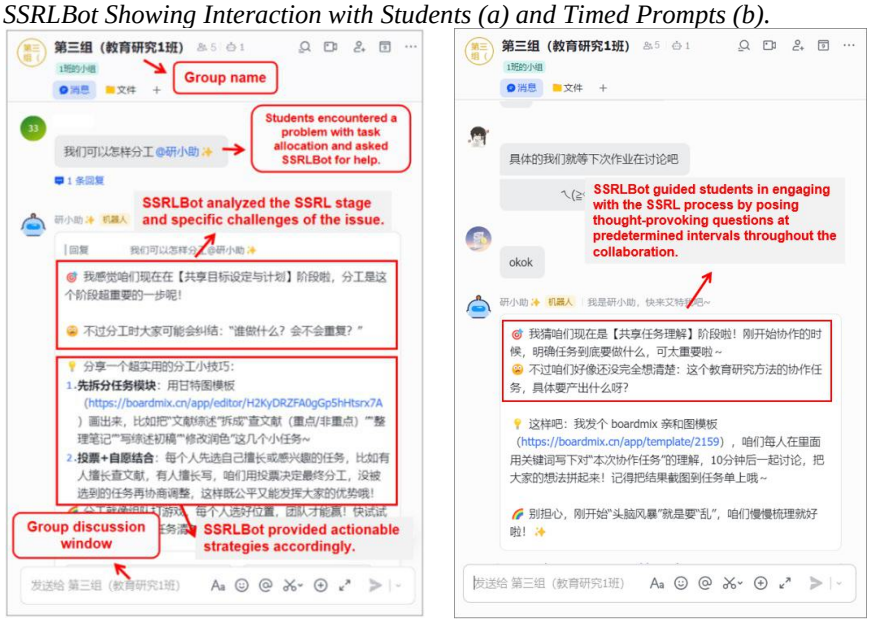
The design of SSRLBot was guided by established SSRL scaffolding principles (Edwards et al., 2025; Järvelä et al., 2015; Kim et al., 2025). These principles emphasize: (a) the importance of enhancing learners' awareness of the entire SSRL process, which includes shared task understanding, shared goal setting and planning, shared monitoring and strategy adjustment, and shared reflection; and (b) the need to support students in applying appropriate strategies at each stage of SSRL to facilitate the regulation process. SSRLBot was developed on the Coze platform and embedded in the Feishu collaborative environment. It was powered by the Doubao large language model and designed as a team moderator rather than a general-purpose conversational assistant. SSRLBot leveraged the semantic understanding capabilities of the LLM, while its responses were constrained by the SSRL framework and supported through Retrieval-Augmented Generation (RAG) with an expert-curated knowledge base of SSRL strategies and course content.

SSRLBot provided two main functions (See Figure 1). First, it delivered proactive, phase-specific prompts at regular intervals to guide groups sequentially through the complete SSRL cycle, explicitly signaling the current regulatory phase and encouraging shared sense-making. Second, it offered on-demand adaptive

support when students queried the agent, diagnosing regulatory challenges and generating strategy-oriented guidance aligned with the group's current SSRL phase, while avoiding direct task answers. A pilot study with five students was conducted prior to deployment to refine the prompting design and ensure system stability.

Figure 1

SSRLBot Showing Interaction with Students (a) and Timed Prompts (b).



(a) (b)

Data collection and analysis

To evaluate the intervention, multidimensional engagement and learning outcomes were assessed. Behavioral engagement was derived from platform logs (i.e., online learning duration and message counts) and participation equality via the Gini coefficient. Cognitive engagement was analyzed by coding posts at the sentence level using the ICAP framework (Chi & Wylie, 2014), with 15% of the data double-coded to ensure reliability ($\kappa = 0.84$). Emotional engagement was measured using a 5-item adapted scale (Xu et al., 2024; $\alpha = 0.891$). Finally, learning outcomes were determined by the quality of group proposals, which were blindly evaluated by two trained raters using a 6-dimension rubric (i.e., completeness, significance, originality and novelty, feasibility, scientificity, and clarity). The maximum possible score per submission was 100 points. The final score was the average of the two raters' scores.

Results and discussion

Due to non-normal distributions and the small group-level sample size ($n = 12$), Mann–Whitney U tests were employed for all comparisons.

Comparison of learning engagement

Behavioral engagement

Regarding behavioral engagement, the SSRLBot-CSCL group generated significantly more discussion messages ($M = 53.37$, $SD = 33.77$) than the T-CSCL group ($M = 32.39$, $SD = 31.75$), $U = 207.00$, $p = 0.004$, $r = 0.39$. Furthermore, Gini coefficient analysis revealed that SSRLBot-CSCL groups exhibited more equal participation ($M = 0.28$, $SD = 0.05$) compared to T-CSCL groups ($M = 0.41$, $SD = 0.11$), $U = 3.00$, $p = 0.016$, $r = 0.69$.

Aligned with Social Interdependence Theory, the chatbot's prompts for shared goal-setting and monitoring likely strengthened collective responsibility and individual accountability, thereby sustaining more balanced active participation (Johnson & Johnson, 1998). While T-CSCL students spent more time viewing materials, the difference was not significant ($p = 0.219$), suggesting the chatbot may shift engagement from individual consumption to interactive, collective problem-solving.

Cognitive engagement

 © ISLS

For cognitive engagement, significant differences emerged across all ICAP categories (Table 1): SSRLBot-CSCL groups contributed more to interactive and constructive discussions, while the T-CSCL groups contributed more to active and passive discussions.

Table 1

Frequency, Percentage , and Mann-Whitney U Tests for Students' Cognitive Engagement.

Type	SSRLBot-CSCL class		T-CSCL class		*U*	*Z*	*p*
	Mean	**Percentage**	**Mean**	**Percentage**			
Passive	9.26	0.17	14.21	0.40	15.000	-5.979	< **0.001*** **
Active	13.11	0.22	15.57	0.40	5.000	-6.157	< **0.001*** **
Constructive	15.74	0.26	5.96	0.13	17.000	-5.945	< **0.001*** **
Interactive	22.30	0.36	3.71	0.07	2.000	-6.213	< **0.001*** **

Note: *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

One theoretical lens for interpreting this pattern is Cognitive Load Theory, which suggests that learners' limited processing capacity can be affected by coordination demands during collaboration. It is possible that the chatbot's guidance for identifying challenges and adjusting joint strategies helped manage these demands, allowing students to allocate more cognitive resources to building on peers' ideas rather than producing passive responses. However, cognitive load was not directly measured in this study, so this interpretation should be considered tentative.

Emotional engagement

For emotional engagement, the Mann-Whitney *U* test indicated that SSRLBot-CSCL groups ($M = 3.95$, $SD = 0.49$) significantly outperformed T-CSCL groups ($M = 3.27$, $SD = 1.03$), $U = 220.50$, $p = 0.007$, $r = 0.36$.

This aligns with Self-Determination Theory: adaptive prompts and actionable strategies can support autonomy, competence, and relatedness—conditions conducive to interest and persistence. Importantly, emotional gains co-occurred with more equitable participation and higher-level discourse, suggesting SSRLBot may support collective coordination rather than merely producing a novelty effect.

Comparison of collaborative learning outcome

For collaborative learning outcome, Mann–Whitney *U* tests indicated that the SSRLBot-CSCL groups ($M = 90.51$, $SD = 1.61$) significantly outperformed the T-CSCL groups ($M = 87.96$, $SD = 1.65$) in total proposal scores, $U = 5.00$, $p = 0.037$, $r = 0.60$. Specifically, significant gains were observed in originality and novelty ($p = 0.035$, $r = 0.61$) and clarity ($p = 0.029$, $r = 0.63$), whereas no significant differences were found in procedural dimensions such as completeness, feasibility or scientificity ($p > 0.05$).

This pattern extends prior work on SSRL and GenAI supports by highlighting the value of a chatbot that orchestrates group regulation rather than acting only as a Q&A tool. The selective effects likely reflect that regulatory scaffolding most directly benefits idea advancement and articulation, whereas procedural quality may require additional domain- and method-specific guidance.

Conclusions and limitations

Prior work has rarely integrated SSRL scaffolding into GenAI chatbots, and existing SSRL-focused chatbot studies remain largely conceptual with limited empirical evaluation. To address this gap, we designed and implemented SSRLBot to diagnose SSRL challenges and provide adaptive strategies during CSCL. Results showed that SSRLBot improved students' behavioral, cognitive, and emotional engagement, as well as collaborative learning outcomes. This study contributes to CSCL in three ways. First, it expands the role of GenAI chatbots beyond the tutor/assistant paradigm by demonstrating a chatbot that functions as a group-level moderator to facilitate SSRL. Second, it provides empirical evidence that SSRL-oriented chatbot scaffolding can enhance both collaborative processes and products. Third, it suggests a practical pathway for supporting learner collective agency and mitigating overreliance on GenAI through regulation-focused guidance.

Several limitations should be acknowledged. While the current study focused on comparing engagement and learning outcomes across conditions, future work could deepen the process-level understanding by examining how students take up and build on specific chatbot prompts and by evaluating the relevance of the guidance over time. Relatedly, beyond coding cognitive engagement levels, subsequent studies may adopt interaction-quality frameworks that also capture how contributions connect to and advance the group's collective work (e.g., Zhang et al., 2009). Finally, to clarify how SSRL-oriented chatbot support translates into engagement and outcomes, larger-sample studies could test explanatory pathways statistically (e.g., mediation models linking regulation, engagement, and performance).

References

Chen, D., Zhang, Y., Luo, H., Zhu, Z., Ma, J., & Lin, Y. (2024). Effects of group awareness support in CSCL on students' learning performance: A three-level meta-analysis. *International Journal of Computer-Supported Collaborative Learning*, *19*(1), 97-129.

Chi, M. T. H., & Wylie, R. (2014). The ICAP Framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, *49*(4), 219–243.

Dara, C., Vann, R., & Sok, V. (2024). The Role of Chatgpt as A Virtual Assistant in Increasing Student Learning Collaboration. *Al-Hijr: Journal of Adulearn World*, *3*(4), 505-516.

Edwards, J., Nguyen, A., Lämsä, J., Sobocinski, M., Whitehead, R., Dang, B., ... & Järvelä, S. (2025). Human-AI collaboration: Designing artificial agents to facilitate socially shared regulation among learners. *British Journal of Educational Technology*, *56*(2), 712-733.

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of educational research*, *74*(1), 59-109.

Hadwin, A. F., Järvelä, S., & Miller, M. (2011). Self-regulated, co-regulated, and socially shared regulation of learning. *Handbook of self-regulation of learning and performance*, *30*, 65-84.

Huang, X., Gao, J., & Wu, H. (2025, July). SSRLBot: Designing and Developing a Large Language Model-Based Agent Using Socially Shared Regulated Learning. *In International Conference on Artificial Intelligence in Education* (pp. 61-68). Cham: Springer Nature Switzerland.

Isohätälä, J., Järvenoja, H., & Järvelä, S. (2017). Socially shared regulation of learning and participation in social interaction in collaborative learning. *International journal of educational research*, *81*, 11-24.

Järvelä, S., Kirschner, P. A., Panadero, E., Malmberg, J., Phielix, C., Jaspers, J., ... & Järvenoja, H. (2015). Enhancing socially shared regulation in collaborative learning groups: Designing for CSCL regulation tools. *Educational Technology Research and Development*, *63*(1), 125-142.

Johnson, D. W., & Johnson, R. T. (1998). Cooperative learning and social interdependence theory. In *Theory and research on small groups* (pp. 9-35). Boston, MA: Springer US.

Kim, J., Detrick, R., Yu, S., Song, Y., Bol, L., & Li, N. (2025). Socially shared regulation of learning and artificial intelligence: Opportunities to support socially shared regulation. *Education and Information Technologies*, 1-39.

Mishra, P., Pandey, C. M., Singh, U., Gupta, A., Sahu, C., & Keshri, A. (2019). Descriptive statistics and normality tests for statistical data. *Annals of cardiac anaesthesia*, *22*(1), 67-72.

Panadero, E., & Järvelä, S. (2015). Socially shared Regulation of learning: A review. *European Psychologist*, *20*(3), 190–203.

Sánchez-Reina, J. R., Gutiérrez-Ferré, A., & Hernández-Leo, D. (2025). Measuring the Effect of GenAI Chatbots in CSCL Environments: A Study With PyramidApp. In Oshima, J., Chen, B., Vogel, F., & Järvelä, J. (Eds.), *Proceedings of the 18th International Conference on Computer-Supported Collaborative Learning - CSCL 2025* (pp. 622-624). International Society of the Learning Sciences.

Sinha, S., Rogat, T. K., Adams-Wiggins, K. R., & Hmelo-Silver, C. E. (2015). Collaborative group engagement in a computer-supported inquiry learning environment. *International Journal of Computer-Supported Collaborative Learning*, *10*(3), 273-307.

Xu, B., Stephens, J. M., & Lee, K. (2024). Assessing student engagement in collaborative learning: development and validation of new measure in China. *The Asia-Pacific Education Researcher*, *33*(2), 395-405.

Zheng, Y., Shang, C., Xu, W., Zhang, P., Zhao, Y., & Liu, Y. (2025). Investigating the role of socially shared regulation of learning in fostering student teachers' online collaborative reflection abilities. *Education and Information Technologies*, *30*(7), 8805-8827.

Zhou, X., & Tsai, C. W. (2023). The effects of socially shared regulation of learning on the computational thinking, motivation, and engagement in collaborative learning by teaching. *Education and Information Technologies*, *28*(7), 8135-8152.

Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist*, *25*(1), 3–17.

Acknowledgments

This study is supported by the Research on the theory and Practice of Computational Thinking oriented Artifcial Intelligence Education Framework in primary and secondary schools (No. 72274076). It is also supported by the Research on the Impact of Integrating Group Awareness and GAI-Supported Collaborative Scripts on Shared Regulation among Undergraduate (No. 2025CXZZ036).

International Society of the Learning Sciences

Regulating Epistemic Emotions in CSCL: Comparing the Higher- and Lower-Performing Groups

Wenli Chen, Nanyang Technological University, Singapore, wenli.chen@nie.edu.sg
Lishan Zheng, Nanyang Technological University, Singapore, nie23.zl0281@e.ntu.edu.sg
Xuanyu Chen, Fudan University, China, xuanyuchen23@m.fudan.edu.cn
Qianru Lyu, Carnegie Mellon University, USA, qlyu@andrew.cmu.edu

Abstract: Epistemic emotions (e.g., curiosity, surprise, confusion) are important indicators of learners' affective engagement in computer-supported collaborative learning (CSCL). While prior work studied learners' static epistemic emotions, this study examined their dynamic, sequential patterns by comparing higher- and lower-performing groups. 44 university students participated in collaborative learning ideation tasks in dyads, where their verbal interactions were collected and coded for epistemic emotions through content analysis and sequential mining. The findings show that high-performing groups manage to channel confusion and curiosity into constructive exploration and affective closure, whereas low-performing groups often fail to resolve such tension, leading to anxiety-driven loops. It highlights the importance of evolving epistemic emotional patterns rather than static affective states to understand and optimize collaborative learning processes and outcomes.

Keywords: Collaborative Learning; Computer-Supported Collaborative Learning; Epistemic Emotions

Introduction

The effectiveness of collaborative learning depends largely on the quality of group members' interdependent interactions (Barron, 2003; Chen, Zheng, et al., 2025). In computer-supported collaborative learning (CSCL), epistemic emotions are affective states (e.g., curiosity, surprise, confusion) that arise from the appraisal of the alignment or misalignment between new information and prior knowledge or beliefs during interactions, which serve as indicators of learning experiences and outcomes, particularly when facing challenging learning situations (Muis et al., 2018). A growing body of research examines how epistemic emotions facilitate or constrain the learning processes and outcomes (Graesser, 2020; Järvelä et al., 2020). For instance, positive emotional states such as curiosity and enjoyment enhance learners' engagement, exploration, and achievement, while negative emotions like boredom and frustration predict learners' disengagement and lower learning performance. However, some studies also indicated that negative emotions (e.g., confusion) may also lead to positive learning outcomes if they are appropriately resolved (Muis et al., 2015). These mixed findings highlight the complexity of the relationship between epistemic emotions and learning performance. Prior research has primarily examined either the prevalence of epistemic emotional states or the links between individual emotions and learning outcomes. However, rather than treating epistemic emotions as static affective states, it is important to understand how they unfold over time during collaborative learning. Accordingly, this study investigates the temporal dynamics of learners' epistemic emotions and how these dynamics shape the persistence, adaptation, or disengagement of higher- and lower-performing groups in CSCL. Specifically, this study addresses the following research questions:
(1) What differences in epistemic emotions are identified between higher- and lower-performing students in CSCL?
(2) What differences in the transitional and sequential patterns of epistemic emotions are exhibited by higher- and lower-performing groups in CSCL?

Literature review

Epistemic emotions and verbal interaction in CSCL

In CSCL environments, emotions are deeply situated within the socio-cognitive dynamics of collaborative inquiry (Pekrun & Perry, 2014). Unlike general academic emotions (Obergriesser & Stoeger, 2020), epistemic emotions are defined as responses directed toward knowledge construction processes, arising when learners appraise the (mis)alignment between new information, prior knowledge, and existing beliefs (Muis et al., 2018; Yang et al., 2024). Common manifestations include curiosity, surprise, enjoyment, confusion, frustration, anxiety, and boredom (Muis et al., 2015), as well as neutral states (Arguel et al., 2019). These internal affective states are externalized through verbal interaction, allowing them to be systematically captured and analyzed.

 © ISLS

Epistemic emotions and learning performance

Epistemic emotions function as critical predictors of collaborative learning performance (Graesser, 2020). Positive epistemic emotions, such as curiosity and enjoyment, are consistently linked to deeper cognitive engagement and superior achievement. Conversely, negative epistemic states, including boredom, frustration, and persistent anxiety, tend to impede cognitive processing and undermine learning outcomes (Muis et al., 2018). In the context of CSCL, Yang et al. (2025) observed distinct emotional profiles based on performance levels, where high-performing groups exhibited significantly more adaptive epistemic emotions (e.g., curiosity and interest), whereas low-performing groups were characterized by maladaptive states (e.g., anxiety, frustration, and boredom). Notably, adaptive emotions, particularly curiosity and well-managed confusion, were found to drive richer knowledge co-construction (Yang et al., 2025). Its impact on learning is determined largely by its resolution. When successfully resolved, confusion acts as a catalyst, sparking renewed curiosity and fostering productive collaborative inquiry. However, when confusion remains unresolved, it risks devolving into frustration and ultimately boredom (Baker et al., 2025), thereby reducing engagement and inhibiting performance.

Sequential patterns of epistemic emotions

Recent research highlights the importance of analyzing epistemic emotions as dynamic, sequential processes that unfold throughout collaborative learning (Hou et al., 2025). Sequential pattern mining (SPM) and lag sequential analysis (LSA) have enabled researchers to detect recurring emotional trajectories, such as joy → curiosity → confusion → joy, which are associated with productive inquiry (Hou et al., 2025). Conversely, trajectories where unresolved confusion escalates into frustration or boredom signal a breakdown in emotional regulation. These sequential patterns diverge significantly based on group performance, which highlights the regulatory function of emotion transitions (Yang et al., 2025). High-performing groups are characterized by their capacity to regulate negative states, frequently converting them into productive ones (e.g., curiosity, enjoyment), whereas low-performing groups tend to remain in or transition between unproductive emotions (Yang et al., 2025) . Even though the presence of epistemic emotions shaping engagement and learning outcomes is well documented, the underlying patterns governing their temporal evolution remain underexplored (Yang et al., 2024). This study traces the sequences and transitions of learners' epistemic emotions while comparing group performance to shed light on the patterns of epistemic emotion that facilitate or hinder productive collaborative learning.

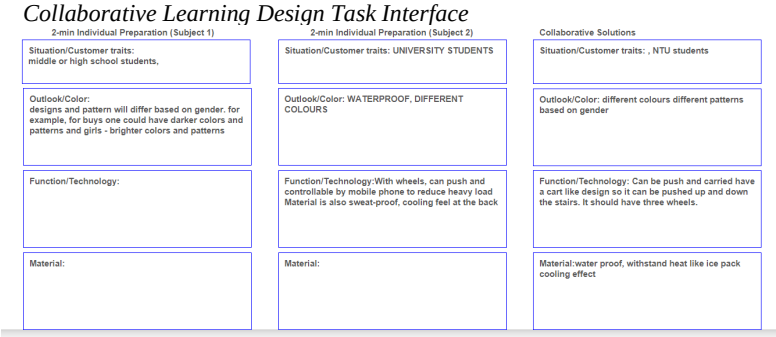
Method

Experimental design

This study employed an experimental design involving 44 participants (22 dyads) recruited from a university in Singapore. The experiment was conducted within a CSCL environment facilitated by a shared online workspace. A crossover design was implemented wherein each dyad completed three comparable product ideation tasks in a randomized order during a single session. Each task lasted for seven minutes. To support knowledge construction, participants were instructed to engage in verbal interaction via Zoom while simultaneously working on a shared Google Slides interface (see Figure 1).

Figure 1

Collaborative Learning Design Task Interface



Data collection and coding framework

All collaborative sessions were recorded via Zoom. All collaborative sessions were recorded via Zoom. Verbal transcripts were generated using the Whisper automatic speech recognition (ASR) system, which produced time-stamped transcripts for each participant. To ensure accuracy, two researchers independently reviewed the machine-generated transcripts against the audio recordings, corrected transcription errors, and reconciled

discrepancies before analysis. The unit of analysis was the turn of talk, defined as a segment beginning when one participant started speaking and ending when another participant took the floor or when a substantial pause occurred. Verbal data were coded using a scheme adapted from established CSCL frameworks (Han et al., 2021; see Table 1). Each turn was categorized into one of eight discrete epistemic states (see Table 1). To establish reliability, two trained coders independently analyzed the full dataset.

Table 1

Epistemic Emotion Coding Framework (Adapted from Han et al. (2021)

Code	Emotion	Description	Example
E1	Confusion	Feeling bewildered or unclear when something cannot be resolved or understood.	"Um, usually what a knee pad or knee guard should do—or rather ideally what they should do—is probably to provide some support, in a sense?"
E2	Frustration	Response to cognitive incongruity from information that conflicts with one's beliefs or prior knowledge.	"I forgot what the material is called. I saw the bag in the catalogue, but I truly forgot."
E3	Enjoyment	Pleasure when information coincides with or verifies one's existing epistemic beliefs.	"Ok, great—great."
E4	Neutral	Indifferent or unclear emotions.	"I'll just add the color again."
E5	Anxiety	Worry, nervousness, or unease, typically about something with an uncertain outcome.	"I completely forgot about function and technology, which I think you covered."
E6	Curiosity	Desire to obtain new knowledge, driven by inconsistency with what one knows.	"So if we were to combine our work, what kind of collaborative idea can we make?"
E7	Surprise	Immediate reaction to cognitive incongruity with prior knowledge or to unexpected information.	"I realized that I was looking completely at the establishment." Yeah, yeah—right, right—I hadn't thought of that. Actually, that lunchbox earlier should probably be like this too focused on the appearance design. Hmm, we might need this."
E8	Boredom	Feeling bored due to something tedious.	"That's fine; we don't really understand what it means anyway."

Findings

RQ1. What are the differences in epistemic emotions identified between higher- and lower-performing groups in CSCL?

To assess collaborative learning outcomes, the design artifacts generated across the three tasks were evaluated using an adapted multidimensional scoring rubric (Chen, Peng, et al., 2025). Performance was quantified on a 5-point scale across four distinct dimensions: Completeness, Elaboration, Functionality, and Innovation. Groups were then divided based on median scores into higher-performing (n=11) and lower-performing (n=11) groups.

Descriptive analysis revealed the presence of eight epistemic emotions across both Higher-Performing Groups and Lower-Performing Groups (see Table 2). In terms of frequency, Neutral states (E4) emerged as the predominant state, accounting for approximately half of all coded turns for both groups. Following this, both groups exhibited substantial levels of Confusion (E1) and Enjoyment (E3). Other emotions, such as Curiosity (E6) and Surprise (E7), appeared less frequently but played roles in triggering exploration. Negative emotions like frustration (E2), anxiety (E5), and boredom (E8) occurred only occasionally. Overall, the frequency distributions indicate that both performance groups experienced emotions, with only slight quantitative differences.

Table 2

Frequency and Proportion of Epistemic Emotions

Epistemic Emotion	Higher Performing Groups	Lower Performing Groups
Confusion（E1）	300 (19.0%)	216 (16.2%)
Frustration（E2）	21 (1.3%)	9 (0.7%)
Enjoyment（E3）	220 (13.9%)	189 (14.2%)
Neutral（E4）	793 (50.2%)	693 (51.9%)
Anxiety（E5）	43 (2.7%)	32 (2.4%)

 © ISLS

Curiosity（E6）	129 (8.2%)	124 (9.3%)
Surprise（E7）	71 (4.5%)	68 (5.1%)
Boredom（E8）	2 (0.1%)	4 (0.3%)

RQ2. What are the differences in transitional and sequential patterns of epistemic emotions exhibited by higher- and lower-performing groups in CSCL?

To investigate epistemic emotion dynamics during collaborative learning, we applied Sequential Pattern Mining (SPM) using the SPADE algorithm in R (Zhang & Paquette, 2023). Following Yang et al. (2022), a minimum support threshold of 0.60 was set to retain only the most common patterns.

The analysis revealed distinct sequential patterns between the groups (Table 3). Higher-performing groups exhibited adaptive trajectories, successfully resolving epistemic tension (curiosity, confusion) into productive or neutral states (e.g., Curiosity → Enjoyment [0.94]). Longer sequences, such as Curiosity → Enjoyment → Neutral, further demonstrated their ability to achieve positive resolution, and anxiety-related sequences were notably absent. Conversely, lower-performing groups displayed maladaptive trajectories where tension frequently escalated into unresolved emotional loops. This was evidenced by patterns like Surprise → Anxiety [0.75] and longer chains such as Curiosity → Anxiety → Surprise. Although lower-performing teams occasionally exhibited positive transitions (e.g., Curiosity → Enjoyment [0.81]), these were less frequent and stable. Ultimately, higher-performing groups effectively regulated epistemic challenges, whereas lower-performing teams frequently became entrapped in anxiety-driven cycles that hindered collaboration.

Table 3

Representative Emotion Subsequences in Higher- And Lower-Performing Groups

Group	Emotion Subsequence	Support Value
Higher Performing	Curiosity → Neutral	1.0000
	Enjoyment → Neutral; Curiosity → Enjoyment; Curiosity → Enjoyment → Neutral	0.9375
	Confusion → Enjoyment; Confusion → Neutral; Curiosity → Confusion; Neutral → Confusion → Enjoyment; Curiosity → Confusion → Enjoyment; Curiosity → Confusion → Neutral; Neutral → Curiosity → Confusion → Enjoyment	0.8750
	Surprise → Neutral; Curiosity → Surprise; Confusion → Surprise	0.7500
Lower Performing	Curiosity → Neutral; Confusion → Neutral; Curiosity → Confusion; Curiosity → Confusion → Neutral	1.0000
	Confusion → Surprise; Surprise → Neutral; Curiosity → Surprise; Surprise → Confusion → Neutral; Curiosity → Confusion → Surprise; Curiosity → Confusion → Neutral; Surprise → Curiosity → Neutral; Surprise → Curiosity → Confusion → Neutral; Confusion → Enjoyment; Enjoyment → Neutral; Curiosity → Enjoyment; Neutral → Confusion → Enjoyment; Curiosity → Confusion → Enjoyment; Curiosity → Enjoyment → Neutral; Neutral → Curiosity → Confusion → Enjoyment	0.8125
	Anxiety → Surprise; Confusion → Anxiety; Anxiety → Neutral; Curiosity → Anxiety; Confusion → Anxiety → Surprise; Surprise → Anxiety → Neutral; Curiosity → Anxiety → Surprise; Confusion → Anxiety → Neutral; Curiosity → Confusion → Anxiety; Curiosity → Anxiety → Neutral; Surprise → Confusion → Anxiety → Neutral; Curiosity → Confusion → Anxiety → Surprise; Surprise → Curiosity → Anxiety → Neutral	0.7500

Conclusion and discussion

This study examined the frequency (RQ1) and temporal trajectories (RQ2) of epistemic emotions in high- and low-performing CSCL groups. While both groups experienced a similarly broad range of emotions, they differed significantly in emotional regulation over time.

High performers demonstrated resolution-oriented trajectories, transforming epistemic tension (e.g., curiosity, confusion) into positive states like enjoyment while completely avoiding anxiety. This aligns with research characterizing resolved confusion as a productive struggle (D'Mello et al., 2014). Conversely, low performers displayed maladaptive trajectories where tension frequently escalated into unresolved anxiety. This finding is consistent with earlier work indicating that unresolved negative emotions can disrupt collaborative learning and constrain collective reasoning (Järvelä et al., 2020). The different epistemic emotions trajectories show that learning performance may be achieved by effectively regulating the negative emotions during the collaborative learning process, underscoring the role of epistemic emotion regulation in knowledge construction and integration.

References

Arguel, A., Lockyer, L., Kennedy, G., Lodge, J. M., & Pachman, M. (2019). Seeking optimal confusion: A review on epistemic emotion management in interactive digital learning environments. *Interactive Learning Environments, 27*(2), 200–210.

Baker, R. S., Cloude, E., Andres, J. M., & Wei, Z. (2025). The confrustion constellation: A new way of looking at confusion and frustration. *Cognitive Science*, *49*(1), e70035.

Barron, B. (2003). When smart groups fail. *The Journal of the Learning Sciences*, *12*(3), 307–359.

Chen, W., Peng, T. W., Caryn, N. E. E., Mavis, H. M. Y., Zheng, L., & Chen, X. (2025). Mental effort and task performance in collaborative learning: The impact of group awareness. *Proceedings of the 18th International Conference on Computer-Supported Collaborative Learning-CSCL 2025, Pp. 116-123*.

Chen, W., Zheng, L., Chen, X., & Ho, M. (2025). How individual preparation before collaboration shapes learners' verbal communication patterns: An epistemic network analysis. *Proceedings of the 19th International Conference of the Learning Sciences-ICLS 2025, Pp. 2855-2857*.

D'Mello, S., Lehman, B., Pekrun, R., & Graesser, A. (2014). Confusion can be beneficial for learning. *Learning and Instruction*, *29*, 153–170.

Graesser, A. C. (2020). Emotions are the experiential glue of learning environments in the 21st century. *Learning and Instruction*, *70*, 101212.

Han, Z.-M., Huang, C.-Q., Yu, J.-H., & Tsai, C.-C. (2021). Identifying patterns of epistemic emotions with respect to interactions in massive online open courses using deep learning and social network analysis. *Computers in Human Behavior*, *122*, 106843.

Hou, C., Zhu, G., Yang, Y., & Tan, S.-C. (2025). Exploring emotional dynamics between productive and improvable knowledge-building discourses. *Computers & Education*, *238*, 105395.

Järvelä, S., Gašević, D., Seppänen, T., Pechenizkiy, M., & Kirschner, P. A. (2020). Bridging learning sciences, machine learning and affective computing for understanding cognition and affect in collaborative learning. *British Journal of Educational Technology*, *51*(6), 2391–2406.

Muis, K. R., Chevrier, M., & Singh, C. A. (2018). The role of epistemic emotions in personal epistemology and self-regulated learning. *Educational Psychologist*, *53*(3), 165–184.

Muis, K. R., Psaradellis, C., Lajoie, S. P., Di Leo, I., & Chevrier, M. (2015). The role of epistemic emotions in mathematics problem solving. *Contemporary Educational Psychology*, *42*, 172–185.

Obergriesser, S., & Stoeger, H. (2020). Students' emotions of enjoyment and boredom and their use of cognitive learning strategies–How do they affect one another? *Learning and Instruction*, *66*, 101285.

Pekrun, R., & Perry, R. P. (2014). Control-value theory of achievement emotions. In *International handbook of emotions in education* (pp. 120–141). Routledge.

Yang, Y., Yuan, K., Zhu, G., & Jiao, L. (2024). Collaborative analytics-enhanced reflective assessment to foster conducive epistemic emotions in knowledge building. *Computers & Education*, *209*, 104950.

Yang, Y., Zhang, Y., Jiang, L., Sun, D., Yuan, W., & Zhu, G. (2025). Importance of epistemic emotions and collaborative interactions in collective knowledge building: Evidence from high-and low-performance groups. *Interactive Learning Environments*, *33*(7), 4309–4333.

Yang, Y., Zhu, G., & Chan, C. K. (2022). Evolution of the academic emotions of academically low-achieving students in knowledge building. *International Journal of Computer-Supported Collaborative Learning*.

Zhang, Y., & Paquette, L. (2023). Sequential pattern mining in educational data: The application context, potential, strengths, and limitations. In *Educational data science: Essentials, approaches, and tendencies: Proactive education based on empirical big data evidence* (pp. 219–254). Springer.

International Society of the Learning Sciences

The Impact of Shared Feedback Visualization on Student Reflection During Simulation Debriefing

Vitaliy Popov, University of Michigan, vipopov@umich.edu
Anouschka van Leeuwen, Utrecht University, a.vanleeuwen@uu.nl

Abstract: This study examined how a cognitive group awareness tool (RADAR) influences medical students' reflective behaviors during online post-simulation debriefing. A between-subjects experiment with 117 medical students compared three conditions: traditional observational debriefing (control), completing a shared decision-making questionnaire before debriefing (light RADAR), and completing the questionnaire with visualized results (full RADAR). Both RADAR conditions increased reflection on decision-making and counseling behaviors. Epistemic network analysis identified three distinct reflection patterns: control students connected performance review with patient history gathering; light RADAR students linked searching for alternative approaches with communication techniques and counseling; full RADAR students connected performance review with patient partnerships and counseling information. Visualization of assessment discrepancies in full RADAR did not increase feed-forward reflection. This study contributes to understanding how cognitive group awareness tools can enhance reflective learning processes in online simulation-based learning.

Introduction

Simulation-based training has become integral to medical education, providing immersive, risk-free scenarios that mirror real-world clinical challenges (Chernikova et al., 2020; Popov et., 2024). Post-simulation debriefing, typically guided by instructors, plays a pivotal role in maximizing learning outcomes by translating simulated experiences into practical knowledge and skills (Kolbe et al., 2023). However, meta-analytical evidence shows inconsistent effects on learning gains due to substantial variation in debriefing quality and implementation challenges (Cheng et al., 2014; Keiser & Arthur, 2021; Sawyer et al., 2016). To address these limitations, there is growing interest in leveraging learning analytics dashboards (Harmer et al., 2026) and group awareness tools (GATs) (Popov et al., 2025). These tools aggregate and visualize performance indicators with the aim of facilitating learning processes and outcomes (Ollesch et al., 2021). GATs provide near real-time, quantifiable feedback on specific metrics while incorporating reference frames that offer learners comparative points (Jivet et al., 2017). The present study investigates how a cognitive GAT called RADAR affects medical students' reflective behaviors during online post-simulation debriefing. While GATs show promise in providing targeted feedback, little empirical research examines how learners cognitively process such feedback during debriefing or how it impacts reflection quality. This study addresses this knowledge gap by comparing traditional debriefing, light RADAR (questionnaire without visualization), and full RADAR (questionnaire with visualization and reflection prompts) conditions.

Theoretical framework

According to Rudolph and colleagues (2007), learners assign meaning to external stimuli through internal cognitive frames (i.e., mental representations of perceived reality) that remain invisible yet are inferable from actions and results. The primary goal of debriefing is to make these frames visible by externalizing the thinking behind actions, analyzing their impact, and refining and/or creating new frames. Learning occurs particularly when instructors and trainees explore the frames-actions-results causal sequence in reverse. Feedback during post-simulation debriefing is defined as the transmission of information regarding actions, events, processes, or behaviors relative to task completion or teamwork (Gabelica et al., 2014; Sawyer et al., 2016), and is recognized as central to both motivation and learning due to its informational value (Hattie & Timperley, 2007).

The cognitive processing of feedback typically involves three reflective stages (Gabelica et al., 2014): reviewing performance by evaluating present and past performance strategies; looking for alternatives by considering different approaches; and looking ahead by deciding what feedback to implement for future encounters. These stages are cyclical and intended to be repeated for continuous improvement. Despite meta-analytical evidence establishing debriefing's essential role (Keiser & Arthur, 2021; Tannenbaum & Cerasoli, 2013), challenges persist in achieving effective debriefing due to variation in quality, consistency, and limited evidence-based guidance (Mamede & Schmidt, 2023). GATs have emerged as a technological solution to enhance post-simulation debriefing by visualizing performance data and supporting structured reflection (Buder et al., 2021; Janssen et al., 2011). These tools often incorporate reference frames enabling temporal comparisons or

 © ISLS

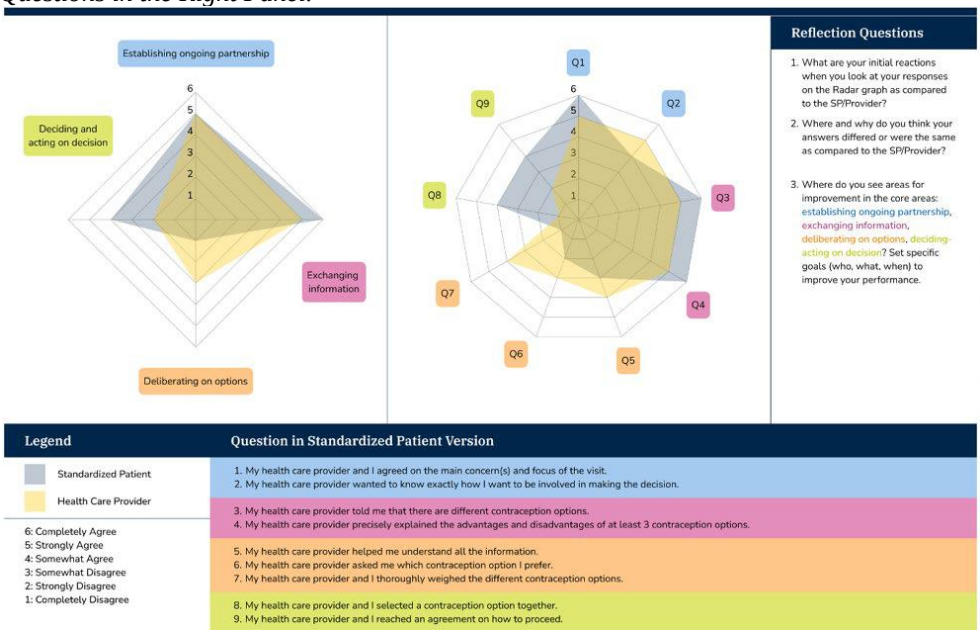
comparisons with others and expectations (Jivet et al., 2017). During debriefing, GATs may visualize students' self-assessments alongside instructor assessments, creating two comparative points that provide insights into previously invisible learner frames and promote metacognitive processes (de Vreugd et al., 2023; Popov, Tan, & Manojlovich, 2024).

Methods

This between-subjects experimental study included 117 medical students (50.4% female, mean age 26 years) enrolled in obstetrics and gynecology clerkships at the University of Michigan Medical School. Students participated in 25-minute online contraception counseling simulations with standardized patients via Zoom, followed by 15-minute debriefing sessions in one of three conditions: traditional observational debriefing (control, n=36), completing a shared decision-making questionnaire before debriefing without viewing scores (light RADAR, n=44), or completing the questionnaire with access to visualized results and guided reflection prompts (full RADAR, n=37). The RADAR tool (Figure 1) mapped four shared decision-making (SDM) sub-constructs (Kriston et al., 2010): establishing ongoing partnership, information exchange, deliberating options, and deciding and acting on decision. Both students and instructors completed the adapted SDM-Q9 questionnaire immediately following simulation (reliability: 0.81 for students, 0.79 for instructors). All debriefing sessions were recorded, transcribed, and segmented based on students' turns of talk. Two independent coders analyzed reflective behaviors using a coding scheme based on Gabelica and colleagues' framework (Table 1), identifying three reflective levels: reviewing performance, looking for alternatives, and looking ahead. Inter-rater reliability was strong (Cohen's kappa = 0.77). Analysis employed non-parametric tests (Kruskal-Wallis and Mann-Whitney U) for frequency comparisons and Epistemic Network Analysis (Shaffer, 2018; Popov & Rochlen, 2025) to examine temporal co-occurrence patterns of codes.

Figure 1

The RADAR Tool, with the Visualization of the Four Dimensions in the Spider Graph, and Reflection Questions in the Right Panel.



Results

Reviewing Performance was the most frequently observed reflective behavior across all conditions (75.5–85.1%), followed by Establishing Partnership (42.9–55.3%), with Looking Ahead occurring rarely (0.5–1.3%). Significant differences emerged across conditions. Students in full RADAR and control conditions reflected more frequently on establishing partnership (55% each) than light RADAR (42.9%, $p<.001$). Control students discussed patient history more frequently (35.2%) compared to full RADAR (20.7%) and light RADAR (15.8%, $p<.001$), while full RADAR students reflected most on counseling behaviors (36.3% versus 21.1% in control and 28.3% in light RADAR, $p<.001$). Both full and light RADAR conditions showed increased reflection on deciding and acting on decision compared to control (11.1% and 9.2% versus 5.5%, $p=.004$). Full RADAR students reviewed overall performance more frequently (85.1% versus 77.8% control and 75.4% light RADAR, $p<.001$), while light

 © ISLS

RADAR students discussed alternatives most often (29.3% versus 17.7% in full RADAR and 16.8% in control, p<.001). No significant differences appeared for domain general communication, deliberating on options, or looking ahead across conditions. Epistemic Network Analysis revealed distinct reflection patterns by condition (Figure 2a,b,c). Control students primarily connected performance review with patient history gathering and establishing partnership. Light RADAR students frequently linked searching for alternatives with communication techniques, patient rapport building, and counseling. Full RADAR students most often connected performance review with establishing patient partnerships and counseling information. Contrary to our expectations, visualization of assessment discrepancies did not increase feed-forward reflection.

Table 1

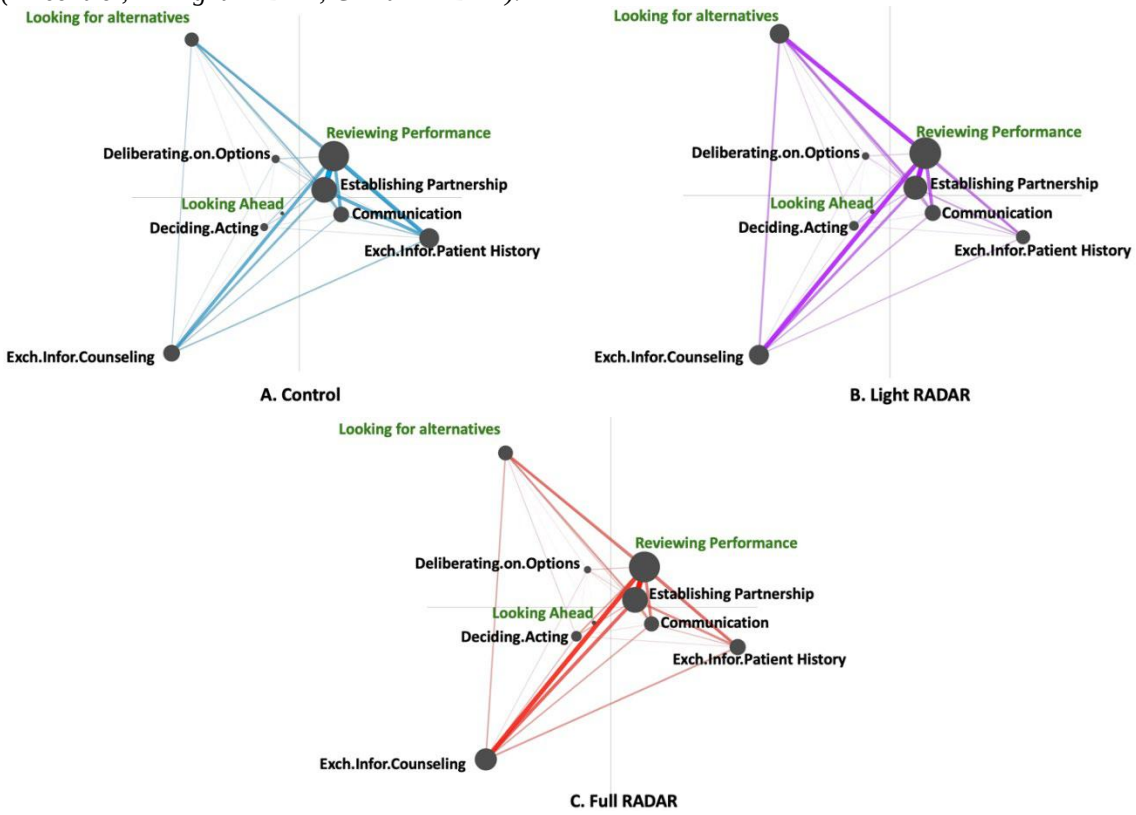
Coding Scheme for Shared Decision-Making Aspects and Reflective Behaviors

Category	Code	Description	Example from data
Shared Decision-Making	Domain general (communication)	Statements related to communication skills and general interpersonal aspects not specific to contraception counseling; includes transitions, time management, and non-verbal communication.	"The transition there wasn't very good, even though I needed that information."
	Establishing partnership	Statements about building trust and mutual respect; assessing patient's preferences for involvement and control over decisions.	"I was good at validating the responses when discussing her pregnancy concerns."
	Exchanging information (patient history)	Statements about eliciting relevant patient history, values, and preferences during the encounter.	"I asked about your goals for pregnancy and past medical history, though I deviated when going through the full list."
	Exchanging information (counseling)	Statements about sharing technical information regarding contraceptive options, benefits, and side effects; communicating research-based evidence.	"I went into detail about the pill but felt I described similar options rather than very different choices."
	Deliberating on options	Statements reflecting on the process of considering and discussing pros and cons of relevant contraceptive choices.	"I could have explicitly asked what was most important to help guide the conversation more clearly."
	Deciding and acting on decision	Statements about helping patients explore preferences and make informed decisions; discussing implementation and next steps.	"I helped you make a decision but didn't follow through with what happens next."
Reflective Behaviors	Reviewing performance	Statements assessing own performance, including goal achievement, strategies used, and reasons for success or failure.	"I kept the interaction fairly organized going through things step by step and was able to elicit the patient's main preferences."
	Looking for alternatives	Statements discussing how the task could have been approached differently; considering alternative sequences of action.	"I wish I would have made more time to do teach-back by having the patient explain it back to me."
	Looking ahead	Statements making explicit decisions about future approaches; stating clear intentions to implement new behaviors in future encounters.	"Next time, I will ask the patient whether they understand the information after I provide new pieces of information."

 © ISLS

Figure 2

ENA Mean Network Models Summarizing Student SDM and Reflective Behaviors (in Green) in Three Conditions (A=control; B=Light RADAR; C=Full RADAR).



Black nodes represent the codes, and thicker, more saturated lines between the nodes represent stronger connections. The larger the node size is the higher frequency, and the code is being used in connection to other nodes. ENA, epistemic network analysis.

Discussion and conclusion

This study contributes to understanding how cognitive group awareness tools enhance reflective learning in online simulation-based education. The results demonstrate that both RADAR conditions shifted reflection toward counseling and deciding and acting on decision compared to traditional debriefing, indicating that structured self-assessment tools direct student attention to specific shared decision-making competencies that may otherwise be overlooked in unstructured debriefing. This attention-directing function represents a key affordance of GATs by structuring assessment around specific competencies, these tools provide cognitive scaffolding that breaks down complex clinical interactions into atomic components for feedback, practice, and remediation. While both RADAR conditions increased reflection on these decision-making domains, they maintained existing reflection areas such as establishing partnership and patient history.

Light RADAR stimulated the highest frequency of higher-order reflection (looking for alternatives), suggesting that questionnaire-based prompts alone offer substantial educational value. The mechanism behind this pattern may relate to how students internalize assessment criteria through questionnaire completion. The cognitive act of self-evaluating performance against specific criteria appears to foster exploration of alternative approaches. In contrast, the full RADAR condition heightened level 1 reflection (reviewing performance) but not higher stages (looking for alternatives and feed-forward reflection), possibly because the visualization of discrepancies between student and instructor assessments creates coordination costs that consume cognitive resources otherwise available for deeper reflection. This finding suggests that visualization complexity, while intended to enhance reflection through explicit discrepancy highlighting, may limit the reflective process when learners must coordinate interpretation of multiple information sources during live debriefing discussion. Our future will examine how individual differences in self-assessment accuracy and the direction of self-assessment bias (overestimation versus underestimation) influence reflection patterns, particularly concerning deliberation and feed-forward planning behaviors.

 © ISLS

References

Buder, J., Schwind, C., Ruping, S., Marquart, C., & Shaikh, M. (2021). Towards learning analytics for personalized learning: Data infrastructure for the classroom. *Computers & Education, 163*, 104088.

Chernikova, O., Heitzmann, N., Stadler, M., Holzberger, D., Seidel, T., & Fischer, F. (2020). Simulation-based learning in higher education: A meta-analysis. *Review of Educational Research, 90*(4), 499–541.

Cheng, A., Morse, K. L., Rudolph, M. I., Button, D., Ritchie, S. D., & Chang, T. P. (2014). Addressing the gap between teaching and clinical practice in pediatric resuscitation. *Academic Medicine, 89*(2), 212–218.

de Vreugd, M., Bastiaens, T. J., & Kirschner, P. A. (2023). Towards determining the role of visual representations in learning analytics dashboards: A systematic literature review. Computers & Education, 194, 104706.

Gabelica, C., Van den Bossche, P., De Maeyer, S., Segers, M., & Gijselaers, W. H. (2014). The effect of team feedback and guided reflexivity on team performance change. *Learning and Instruction, 34*, 86–96.

Harmer, B., Erickson, J., Ross, S., Bhatnagar, D., Dowker, S., Park, H., Sample, A. P., Hartley, S., Cooke, J., Cole, M., & Popov, V. (2026). Development and Evaluation of a Multimodal Debriefing Dashboard for Virtual Reality Cardiac Arrest Simulation. *Simulation in Healthcare*, *21*(1), 48–56.

Hattie, J., & Timperley, H. (2007). The power of feedback. Review of Educational Research, 77(1), 81–112.

Janssen, J., & Kirschner, P. A. (2020). Collaborative learning as a way to improve student engagement? A closer look. *Technology, Pedagogy and Education, 29*(2), 177–186.

Janssen, J., Erkens, G., Kanselaar, G., & Jaspers, J. (2011). Visualization of agreement and discussion processes during computer-supported collaborative learning. *Computers & Education*, *36*(4), 351–366.

Jivet, I., Scheffel, M., Specht, M., & Drachsler, H. (2017). License to monitor: Examining information use of learning dashboards. In Proceedings of the 7th Learning & Analytics & Knowledge Conference (pp. 5–14).

Keiser, H. N., & Arthur, W. (2021). A meta-analytic review of the instructional effectiveness of computer-based simulation games. *Computers & Education, 165*, 104139.

Kolbe, M., Eppich, W., & Rudolph, M. I. (2023). Managing psychological safety in healthcare teams. *Academic Medicine, 98*(2), 159–165.

Kriston, L., Scholl, I., Hölzel, L., Simon, D., Loh, A., & Härter, M. (2010). The 9-item Shared Decision Making Questionnaire (SDM-Q-9). Development and psychometric properties in a primary care sample. *Patient Education and Counseling*, *80*(1), 94–99.

Mamede, S., & Schmidt, H. G. (2023). Innovations in problem-based learning: New approaches to medical education. *Medical Teacher*, *45*(3), 229–237.

Ollesch, G., Preuß, S., Meier, A., & Roßnagel, A. (2021). Group awareness tools—A multidisciplinary review. Computer Supported Cooperative Work (CSCW), 30(1), 1–43.

Popov, V., Mateju, N., Jeske, C., & Lewis, K. O. (2024). Metaverse-based simulation: a scoping review of charting medical education over the last two decades in the lens of the 'marvelous medical education machine'. *Annals of medicine*, *56*(1), 2424450.

Popov, V., Nguyen, S., & Ochoa, X. (2026). Applying multimodal learning analytics to naturalistic recordings of clinical simulations: towards an accurate and scalable pipeline for automated feedback generation. *Learning and Instruction*, *102*, 102267.

Popov, V., Gabelica, C., Tomaka, S., & Danciu, T. (2025). Making the invisible visible: how multisource feedback and guided facilitation affect team reflection. *Cognition, Technology & Work*, 1-20.

Popov, V., & Rochlen, L. R. (2025). From communication to action: using ordered network analysis to model team performance in clinical simulation. *BMC medical education*, *25*(1), 479.

Popov, V., Tan, Y., & Manojlovich, M. (2024). Applying ordered network analysis to video-recorded physician–nurse interactions to examine communication patterns associated with shared understanding in inpatient oncology care settings. *BMJ open*, *14*(6), e084653.

Rudolph, J. W., Simon, R., Rivard, P., Dufresne, R. L., & Raemer, D. B. (2007). Debriefing with good judgment: Combining rigorous feedback with genuine inquiry. *Anesthesiology Clinics*, *25*(2), 361–376.

Sawyer, T., Eppich, W., Brett-Fleegler, M., Grant, V., & Cheng, A. (2016). More than one way to debrief: A critical review of healthcare simulation debriefing practices. *Simulation in Healthcare, 11*(3), 209–217.

Shaffer, D. W. (2018). Quantitative ethnography. Cathcart Press.

Tannenbaum, S. I., & Cerasoli, C. P. (2013). Do team and individual debriefings enhance performance? A meta-analysis. *Human Factors, 55*(1), 231–245.

Acknowledgments

This material is based upon work supported by the U.S. National Science Foundation under Grants No. IIS-2202451 and DRL-2506865.

International Society of the Learning Sciences

Exploring Unstructured Collaboration While Navigating Misinformation in Computer-Supported Collaborative Argumentation: Implications for Design

Ram Das Rai, Sahana Murthy
ramdas97@iitb.ac.in, sahanamurthy@iitb.ac.in
Indian Institute of Technology Bombay

Abstract: The proliferation of socio-scientific misinformation is an inherently social challenge. While Computer-Supported Collaborative Argumentation (CSCA) environments offer a promising pedagogical approach for collective knowledge construction, their effectiveness relies heavily on productive group dynamics. To avoid over-scripting future interventions, this formative Design-Based Research (DBR) study explores the baseline collaborative breakdowns that occur when students navigate misinformation without explicit social guidance. The study analyzed the audio transcripts of two middle school triads collaboratively mapping arguments about a controversial topic, utilizing established interaction analysis frameworks to evaluate peer talk and social modes of co-construction. Findings indicate that, within these groups, students defaulted to two distinct dysfunctional patterns: an epistemic conflict avoidance pattern, characterized by cumulative talk and quick consensus-building, and a social dominance pattern, characterized by disputational, competitive fact-checking. Structural scaffolds alone appeared insufficient to foster the critical exploratory talk required to evaluate deceptive claims, absent social interaction protocols. The study concludes by offering targeted design implications, suggesting that future CSCA environments might benefit from embedded collaboration scripts, such as structured "Devil's Advocate" and "Lead Fact-Checker" roles, to disrupt these natural tendencies and better support apt epistemic performance.

Introduction and theoretical background

The proliferation of misinformation surrounding socio-scientific issues presents a societal challenge with the potential to undermine informed decision-making (McBrayer, 2020). In what has been characterized as a "post-truth" condition, educational responses have frequently relied on individualistic media literacy approaches, focusing on teaching students isolated fact-checking skills (Barzilai & Chinn, 2020). Yet such approaches may struggle to address the reality that misinformation rarely exists in a vacuum; rather, it thrives within social networks where echo chambers, socio-affective factors, and collective validation allow false claims to take root and persist (Ecker et al., 2022). Navigating this landscape may therefore require moving beyond individual fact-checking toward cultivating "apt epistemic performance", the capacity to achieve epistemic aims through reliable processes across diverse contexts, an ability that involves learners participating in epistemic evaluations together with others (Chinn, Barzilai, & Duncan, 2020). Because the spread and acceptance of misinformation are inherently social phenomena, educational interventions designed to counteract them could benefit from being socially embedded (Mercier & Sperber, 2017).

Computer-Supported Collaborative Learning (CSCL) offers a relevant paradigm for addressing this challenge by shifting instructional focus from individual cognition toward collective knowledge construction. Within this domain, Computer-Supported Collaborative Argumentation (CSCA) employs digital environments to foster evidence-based dialogue, encouraging learners to jointly articulate claims, justify them with data, and evaluate competing perspectives (Andriessen & Baker, 2013). Recent advances in digital epistemic scaffolds suggest that explicitly visualizing multiple documents and their sources through shared argument maps can help improve students' abilities to critically evaluate and integrate conflicting information (Barzilai, Mor-Hagani, Zohar, Shlomi-Elooz, & Ben-Yishai, 2020). By rendering reasoning visible and encouraging participants to draw on a shared corpus of information, these environments hold potential for building resilience against false narratives.

Despite this theoretical alignment, simply placing students in a CSCA environment with a digital tool does not necessarily ensure productive argumentation. The educational quality of peer discourse can vary considerably based on underlying group dynamics. Wegerif and Mercer (1997) distinguish between three types of peer talk: "disputational talk," characterized by competitive assertions and individualized decision-making; "cumulative talk," where students build uncritically on each other's ideas to maintain social harmony; and "exploratory talk," where ideas are critically yet constructively challenged and justified. Similarly, Weinberger

 © ISLS

and Fischer (2006) observe that learners frequently default to "quick consensus building", accepting contributions without critical modification, or "conflict-oriented consensus building", replacing a partner's perspective, rather than engaging in integration-oriented perspective-taking. When students are tasked with evaluating misinformation collaboratively, their existing cognitive biases, social hierarchies, and conversational habits can shape the trajectory of their learning (Ecker et al., 2022). Without structural support, students may struggle to achieve the exploratory talk conducive to robust misinformation literacy.

To mitigate such discourse challenges, CSCL research has highlighted the importance of scaffolding and scripting (Kollar, Wecker, & Fischer, 2018). While macro-scripts can structure the overall sequence of classroom activities, specific micro-scripts and structured role assignments may help balance participation and guide learners toward critical, evidence-based peer feedback. However, a tension exists within the design of these environments: although scripting is often recommended to guide peer discourse, over-scripting risks stifling natural collaboration, limiting student agency, and constraining organic meaning-making (Kollar et al., 2018). Before designing complex CSCA interventions, therefore, it seems useful to establish a baseline understanding of what students naturally do when confronted with misinformation in a group setting.

This study adopts a formative, Phase 1 Design-Based Research (DBR) approach to explore this space (Puntambekar, 2018). By intentionally removing micro-scripts, this initial cycle aims to observe middle school students' natural collaborative tendencies when facing the cognitive and socio-affective demands of misinformation. Exploring such unstructured collaboration may help identify where and how future CSCA designs might provide targeted scaffolding. To that end, the primary empirical objective focuses on observable phenomena within student groups, asking: (1) What specific collaborative breakdowns occur when middle school students navigate socio-scientific misinformation in an unstructured CSCA environment? Subsequently, as a secondary theoretical synthesis of these findings, the study asks: (2) What are the design implications of these observed breakdowns for structuring targeted digital scripts and roles in future CSCA environments? The following section describes the methodology employed to investigate these questions, after which the findings and their implications for design are presented.

Methodology

To investigate the research questions, this study adopted a qualitative, formative Design-Based Research (DBR) methodology. In its initial cycle, DBR often involves observing learners in relatively naturalistic settings prior to introducing complex interventions, allowing researchers to gather foundational insights that can inform subsequent design iterations (Puntambekar, 2018). The study was therefore designed to observe how students naturally interact when asked to evaluate and map arguments regarding a polarizing socio-scientific issue: the efficacy of homeopathy. It was selected for its cultural familiarity to participants yet considerable surrounding digital misinformation and conflicting claims (Ernst, 2016). The pilot study included six secondary school students aged 11 to 15 from middle-class backgrounds in Mumbai, India, recruited through convenience sampling. To facilitate observation of small-group dynamics, participants were divided into two triads: Team 1 (Anika, Ananya, Mitali) and Team 2 (Shivam, Kriti, Anuj).

The pedagogical intervention took place during a single-session workshop. The procedure began with a brief assessment of students' baseline beliefs, followed by a tutorial on basic argumentation principles and the digital mapping platform, Miro. Students then viewed an animated scenario depicting a two-person debate about homeopathy between characters Riya and Priya. They were provided with internet access and tasked with collaboratively constructing an argument map to reach a final group verdict on the issue. Students received a visual argument template loosely based on Toulmin's framework to structure claims and evidence (Toulmin, 1958), providing structural scaffolding. However, their social interactions and group roles were intentionally left unstructured. This absence of social scripting was implemented to help surface students' baseline epistemic practices and natural collaborative tendencies when confronting misinformation. The researchers functioned as observers and facilitators, with the reflexive decision to minimize real-time intervention intended to allow authentic peer dynamics to emerge rather than eliciting teacher-pleasing performances.

The primary data sources consisted of continuous audio recordings of the two triads during their collaborative argument-mapping phase, transcribed verbatim to capture nuances of peer discourse including overlaps, interruptions, and off-task commentary. To systematically analyze the transcripts, a deductive coding framework (Table 1) was synthesized from foundational literatures on Computer-Supported Collaborative Learning and misinformation. The interaction analysis applied three distinct theoretical lenses. First, the quality of peer talk was evaluated using Wegerif and Mercer's (1997) dialogical framework, categorizing utterances into exploratory talk, disputational talk, and cumulative talk. Second, to understand how knowledge was negotiated, the analysis utilized Weinberger and Fischer's (2006) social modes of co-construction, coding for interactions such as quick consensus building, conflict-oriented consensus building, externalization, and elicitation. Third,

students' engagement with informational content was coded using a misinformation evaluation framework, tracking instances of source evaluation, fact-checking, and recognition of cognitive biases or barriers to belief revision (Barzilai, Mor-Hagani, Zohar, Shlomi-Elooz, & Ben-Yishai, 2020; Ecker et al., 2022).

Table 1
Codebook for Analysis (Partial)

Category	Code	Description	Source
Peer Talk	Exploratory Talk	Partners engage critically but constructively with each other's ideas.	Wegerif, R., & Mercer, N. (1997)
Epistemic Dimension	Non-Epistemic Activities	Utterances or coordinating moves that have no explicit reference to the learning material or the task of solving the problem.	Weinberger, A., & Fischer, F. (2006)
Social Modes of Co-Construction	Quick Consensus Building	Accepting the contributions of learning partners without modification in order to move on with the task, rather than out of true conceptual agreement.	(Weinberger & Fischer, 2006)
Misinformation Evaluation	Source Evaluation	The process of assessing the credibility, expertise, and trustworthiness of information sources.	Ecker et al. (2022); Barzilai et al. (2020)

Additional relevant codes that informed the analysis can be accessed here: https://docs.google.com/spreadsheets/d/1tZw5znJ72lxhmB2qq6HfwLDjOPc_jL1Oo34iwzT93Ew/edit?usp=sharing

Findings

The interaction analysis of the unstructured collaborative phase revealed distinct patterns in how the two triads navigated the socio-scientific misinformation task. While both groups utilized the digital argument-mapping template, their underlying social dynamics appeared to influence their engagement with the task's epistemic requirements in markedly different ways.

Team 1: Epistemic conflict avoidance and cumulative consensus

Early in the task, an unequal distribution of cognitive labor emerged in Team 1. One student, Anika, explicitly requested collaboration, asking, "Mujhe kyun, tum dono bhi to bolo" (Why me, you both should also speak). When another member, Mitali, opted out by replying, "I don't know," Anika assumed the role of content director while Mitali acted primarily as a scribe. As the session progressed under time constraints, the team appeared to prioritize task completion over complex reasoning. When finalizing their verdict, they chose a stance based on vague impressions of the characters' arguments, with Anika conceding, "Huh... yeah! She had better points!" During a moment when Anika attempted to articulate a complex thought involving the word "argumentative," Mitali, focused only on typing, asked for the spelling. When Anika expressed frustration, Mitali dismissed the task's value entirely, stating, "No one cares! Just go ahead!"

Team 1's discourse predominantly reflected cumulative talk (Wegerif & Mercer, 1997), characterized by uncritical acceptance of each other's ideas in service of maintaining group harmony. Rather than interrogating the scientific plausibility of information, the team engaged in what might be described as epistemic conflict avoidance. Their collaborative process relied heavily on quick consensus building (Weinberger & Fischer, 2006), accepting contributions without substantive modification primarily to move the task forward. The dismissal of reasoning in favor of finishing quickly suggested a reliance on what Weinberger and Fischer (2006) term non-epistemic activities, such as focusing on spelling or making offhand comments, to cope with the task's cognitive demands. By bypassing critical evaluation, the team reached consensus without engaging in the source evaluation necessary for robust misinformation literacy (Ecker et al., 2022).

Team 2: Social dominance and disputational fact-checking

In contrast, Team 2's interactions were characterized by interpersonal friction and highly asymmetric participation. One member, Shivam, largely monopolized both the digital workspace and the conversation. During a debate over the cost of homeopathic remedies, another student, Anuj, interrupted to argue that conventional medicines are "more cheaper." In response, Shivam immediately performed a web search and read from his screen: "There is no significant difference between the two," retorting, "It's the same price!" The social hierarchy driving these interactions was made explicit later when Shivam dismissed his peer's input, stating, "I am smarter than you... You only said yourself." This dynamic appeared to marginalize the third member, Kriti, who remained mostly silent, and prompted Anuj to resort to off-task behavior.

Team 2's interactions aligned closely with disputational talk (Wegerif & Mercer, 1997), wherein students appeared to treat the collaborative activity as a competitive arena rather than a shared epistemic endeavor. When disagreements arose, the group defaulted to conflict-oriented consensus building (Weinberger & Fischer, 2006), resolving issues primarily through social dominance rather than reasoned negotiation. This hierarchical dynamic fundamentally shaped the act of fact-checking. When Shivam conducted his web search, it served as a reactive measure to win an argument rather than as a tool for shared critical inquiry. The team did not engage in source evaluation to determine the credibility of the information (Barzilai et al., 2020), suggesting that the disputational dynamic short-circuited the collaborative sensemaking process.

Synthesis of observed patterns

Addressing the first research question, the observations suggested a bifurcation in group dynamics that appeared to hinder critical inquiry when students navigated misinformation in an unstructured CSCA environment. In these two cases, students tended to default either to cumulative talk (Team 1), which led to epistemic conflict avoidance, or to disputational talk (Team 2), which resolved issues through social dominance. In both scenarios, the absence of structured interaction protocols appeared to constrain the emergence of exploratory talk (Wegerif & Mercer, 1997). Consequently, neither group meaningfully engaged in the deep source evaluation or reflective fact-checking required to navigate deceptive claims. These patterns, while observed in only two triads, offered preliminary insight into potential breakdown points that might inform design considerations, which the following section explores.

Discussion and conclusion

The findings suggest that unstructured CSCA environments may not inherently foster the critical discourse needed to evaluate socio-scientific misinformation. Students defaulted to either cumulative talk, epistemic conflict avoidance and quick consensus building (Team 1), or disputational talk, social dominance and conflict-oriented consensus building (Team 2). Neither group engaged in the exploratory talk or systematic source evaluation foundational to misinformation literacy (Wegerif & Mercer, 1997; Barzilai et al., 2020). These observations align with Weinberger and Fischer's (2006) claim that learners rarely achieve integration-oriented consensus without explicit guidance, extending that insight to the context of misinformation, where cognitive and socio-affective demands are high (Ecker et al., 2022). The results suggest that structural scaffolds, in absence of explicit social protocols, do not reliably regulate the group dynamics that shape engagement with deceptive claims. This points to a need for targeted social scripting (Kollar, Wecker, & Fischer, 2018). To disrupt quick consensus building, future designs could incorporate a structured "Devil's Advocate" role that formally assigns responsibility for challenging emerging consensus. To counter weaponized fact-checking and social dominance, a "Lead Fact-Checker" role paired with turn-taking prompts might transform source evaluation into a shared epistemic responsibility. These proposed scripts aim to balance participation and foster the exploratory talk required for apt epistemic performance (Chinn, Barzilai, & Duncan, 2020).

Methodologically, the decision to minimize researcher intervention proved valuable for surfacing authentic peer dynamics; the need for embedded scripts might have remained obscured had facilitators constantly intervened. The insights, grounded in a specific socio-scientific topic and a small sample of Indian middle-school students, may hold transferability to other collaborative settings where polarizing topics and adolescent social hierarchies intersect, though the sample size (six students across two triads) and single-session design limit claims about generalizability or long-term change. In summary, unstructured collaboration appears to devolve into either uncritical harmony or adversarial dominance, both impeding critical evaluation. CSCA environments designed to counter misinformation may therefore benefit from embedding structured roles and interaction protocols that gently guide students toward exploratory talk and equitable critique. Future research could usefully test such script-enhanced designs across larger, more diverse cohorts and over longer timeframes.

 © ISLS

References

Andriessen, J., Baker, M., & Suthers, D. D. (2013). Argumentation, computer support, and the educational context of confronting cognitions. In J. Andriessen, M. Baker, & D. D. Suthers (Eds.), Arguing to Learn: Confronting Cognitions in Computer-Supported Collaborative Learning Environments (pp. 1–26). Springer Science & Business Media.

Barzilai, S., & Chinn, C. A. (2020). A review of educational responses to the "post-truth" condition: Four lenses on "post-truth" problems. Educational Psychologist, 55(3), 107–119.

Barzilai, S., Mor-Hagani, S., Zohar, A. R., Shlomi-Elooz, T., & Ben-Yishai, R. (2020). Making sources visible: Promoting multiple document literacy with digital epistemic scaffolds. Computers & Education, 157, 103980.

Chinn, C. A., Barzilai, S., & Duncan, R. G. (2020). Education for a "post-truth" world: New directions for research and practice. Educational Researcher, 50(1), 51–60.

Ecker, U. K., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L. K., Brashier, N., ... & Amazeen, M. A. (2022). The psychological drivers of misinformation belief and its resistance to correction. Nature Reviews Psychology, 1(1), 13-29.

Ernst, E. (2016). Homeopathy-the undiluted facts: Including a comprehensive AZ lexicon. Springer.

Kollar, I., Wecker, C., & Fischer, F. (2018). Scaffolding and Scripting (Computer-Supported) Collaborative Learning. In F. Fischer, C. E. Hmelo-Silver, S. R. Goldman, & P. Reimann (Eds.), International Handbook of the Learning Sciences (pp. 340–350).

McBrayer, J. P. (2020). Beyond fake news: Finding the Truth in a World of Misinformation (1st ed.). Routledge.

Mercier, H., & Sperber, D. (2017). The enigma of reason. Harvard University Press.

Puntambekar, S. (2018). Design-Based Research (DBR). In F. Fischer, C. E. Hmelo-Silver, S. R. Goldman, & P. Reimann (Eds.), International handbook of the learning sciences (pp. 383–392). Routledge.

Toulmin, S. E. (1958). The uses of argument. Cambridge University Press.

Wegerif, R., & Mercer, N. (1997). A dialogical framework for researching peer talk. In Computers and talk in the primary classroom (pp. 49–64). Multilingual Matters.

Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. Computers & education, 46(1), 71-95.

International Society of the Learning Sciences

Multi-Level Knowledge Building Engagement: Pathways to Collaborative Problem-Solving and Academic Excellence in Higher Education

Chih Hui Seet, National Chengchi University, 111152518@g.nccu.edu.tw
Chih-Hsuan Chang, National Kaohsiung University of Science and Technology, chsuchang@gmail.com
Huang-Yao Hong, National Chengchi University, hyhong@nccu.edu.tw

Abstract: Collaborative problem-solving (CPS) is increasingly vital for 21st-century learning, yet research on innovative pedagogy for cultivating these competencies in higher education remains limited. This study evaluated knowledge building (KB) as an innovative pedagogy to foster CPS in a university research methods course with 46 students. By analyzing actual online discourse on Knowledge Forum (KF), the study examined five levels of students' KB engagement—behavioral, cognitive, thematic, methodological, and product-based—in relation to their academic achievement. Results revealed high levels of active participation, with 54.10% of contributions involving higher-order cognitive activities. Furthermore, multi-level analysis of KB engagement significantly predicted 43.6% of the variance in academic achievement. These findings provide empirical evidence that the KB environment on KF facilitates the practice of CPS while enhancing academic excellence, offering a validated framework for university curriculum innovation.

Introduction

In the contemporary knowledge economy, the cultivation of core competencies such as Collaborative Problem Solving (CPS) has become a vital driver of national development and academic success (Tan et al., 2021; OECD, 2017). While higher education research methods courses are traditionally conducted in isolation, Knowledge Building (KB) pedagogy provides a transformative framework that shifts the focus from individual information acquisition to the collective advancement and iterative improvement of ideas (Bereiter & Scardamalia, 2014). Despite KB's potential to foster higher-order cognitive engagement, its application in university research methods remains underexplored, with limited evidence on how specific online discourse quality and multi-level engagement relate to academic outcomes. To address these research gaps, this study investigates the impact of KB pedagogy through the following research questions:

- RQ1. How do students participate in KB discourse during the course, and what is the effectiveness and quality of their contributions?
- RQ2. To what extent does students' multi-level engagement in online KF discussions predict their academic performance?

Literature review

KB engagement and CPS

In the contemporary knowledge economy, education must shift from individualistic information transmission toward collective knowledge creation through Knowledge Building (KB). This pedagogy is fundamentally oriented toward the collaborative generation and "idea improvement," a central principle emphasizing that all ideas are improvable through continuous iterative processes that align classroom inquiry with professional knowledge work (Scardamalia & Bereiter, 2003, 2006). Such collective progress relies on Knowledge Building Engagement, which is conceptually defined as an intellectually energized focus and deep commitment to the community's problem-solving and idea-refinement processes. Scardamalia and Bereiter (2021) emphasize that this shift from individual learning to community-level knowledge advancement is a critical transformation that promotes both innovation and students' academic performance.

CPS and online discourse

Collaborative problem solving (CPS) is a vital 21st-century competence involving the capacity to effectively participate in processes where multiple agents work together by sharing understanding and pooling skills toward a solution (OECD, 2017). Beyond a simple skill set, CPS is a complex socio-cognitive competence encompassing dimensions such as social regulation, perspective-taking, and collective knowledge building (Hesse et al., 2015; Chen et al., 2019). Knowledge Building (KB) is recognized as an optimal environment for CPS development

 © ISLS

because its emphasis on collective cognitive responsibility and iterative dialogue aligns with the socio-cognitive demands of collaborative work (Scardamalia, 2002). Online environments, particularly platforms like Knowledge Forum (KF), further facilitate these interactions by providing digital spaces for continuous idea improvement and community knowledge construction (Scardamalia & Bereiter, 2006; Soliman et al., 2021).

Research further underscores the link between high-quality online discourse and academic success, with studies showing that both participation volume and higher-level discourse moves—such as synthesizing and integrating ideas—are significantly correlated with improved domain knowledge (Tong & Chan, 2023). Additionally, the quality of interaction and feedback received in online discussions has been shown to be positively correlated with academic achievement (Jiang et al., 2023). Meta-analytic evidence confirms that online collaborative learning has an overall positive effect on learning performance (Tian & Zheng, 2024). By engaging in iterative processes of idea improvement on KF, students not only deepen their subject-matter expertise but also cultivate the essential collaborative skills required for modern professional knowledge work.

The present study

The present study addresses these critical research gaps. First, while Knowledge Building (KB) is established in K-12 settings, its validation in higher education research methods courses—which naturally require collective inquiry—remains limited (Chen & Hong, 2016; Zhang et al., 2011). Second, while general platform participation is linked to achievement, the predictive influence of qualitative discourse content on academic success necessitates deeper study (Tong & Chan, 2023). Therefore, this study examines the impact of KB pedagogy in a university context to clarify how multi-level online engagement patterns facilitate collaborative skills and academic excellence.

Research Method

Research design, participants, and context

A mixed-methods design was employed to examine students' multi-level learning engagement. Participants (N=46) were second- and third-year undergraduate students, primarily from the Department of Education, enrolled in a compulsory Educational Research Methods course (3 credits) as preparation for education-related graduate programs. The 18-week course combined in-class collaborative activities with online interactions via KF, a computer-supported collaborative KB platform. The course fostered an open, supportive environment emphasizing class discussions and knowledge sharing without rigid structures, aiming to develop a collaborative KB community among students with no prior background in educational research methods.

Course design

The course was designed to integrate collaborative KB pedagogy into this Educational Research Methods course and it was structured in two main parts:

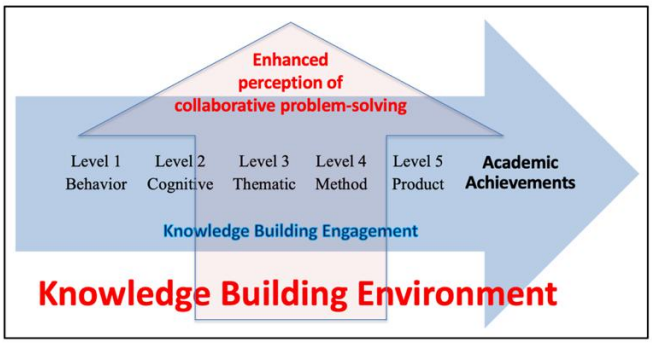
Part 1--Theoretical Instruction. In one part, the instructor delivered lectures, performed regular quizzes (10 times in total), assigning students assignments, and conducted Q&A sessions. Students were introduced to essential concepts in the educational research textbook, including such as research problem formulation, literature review, research design, data collection, qualitative and quantitative analysis, and academic writing etc. Ten class quizzes were administered after the teaching of each essential concept.

Part 2--Collaborative Knowledge Building: In this part, students were required to collectively work on their research proposals from scratch, by translating the conceptual knowledge learned in Part 1 into practical, feasible proposals. Collaboration was community-based and idea-centered, meaning students can interact with any idea or any classmate during online discussion as long as they are interested in the ideas posted in the form of notes on KF through iterative idea interactions to improve their research proposal ideas. As illustrated in Figure 1, the course created a Knowledge Building (KB) Environment where students engaged in idea-centered discussions on the Knowledge Forum (KF). This KB environment guided students through multiple levels of Knowledge Building Engagement (KBE) to foster collective discussion and knowledge advancement, ultimately leading to enhanced perceptions of Collaborative Problem-Solving (CPS) and fostering the development of Academic Achievements in this course of "Educational Research Methods." This KB engagement is operationally defined by five ascending levels of student participation: (1) Level 1: Behavioral Engagement refers to students' online interactions as reflected in their quantitative behaviors on the KF, as reflected in the total number of words generated in KF as derived from specifically note-contributing, note-reading, and note-modifying activities; (2) Level 2: Cognitive Engagement relates to the complexity of students' online contributions, which are evaluated using the six Bloom's cognitive activities: remember, understand, apply, analyze, evaluate, and create; (3) Level 3: Thematic Engagement focuses on whether the students' online discussed topics for their research proposals

show a focus on "future learning competencies" such as team-working, collaborating, and communicating; (4) Level 4: Method Engagement assesses whether students' online discussion is highly related to the textbook content about "educational research methods" in this course, indicating a deep engagement with the course's subject matter; (5) Level 5: Product Engagement refers to the quality of the students' final draft research proposals, serving as the culminating artifact of their collective knowledge advancement efforts. The successful navigation of these engagement levels in the KB environment is theorized to foster both students' CPS perceptions and their overall academic achievements in the course.

Figure 1

Conceptual Framework of the KB Course Design



Data collection and analysis

Data were collected from multiple sources and analyzed to address the research questions. The primary data source was the interactions and discussion content in KF. Students' KF contributions were analyzed across five levels using keywords from KF notes as the unit of analysis: (1) Level 1 (behavior engagement): Descriptive statistics examined students' foundational participation through three activities: note-contributing (creating content), note-reading (understanding content), and note-modifying (enhancing quality), using total word count produced as a predictor. (2) Level 2 (cognitive engagement): Bloom's taxonomy of six cognitive levels—remembering, understanding, applying, analyzing, evaluating, and creating—was used to classify students' cognitive engagement as either lower-order or higher-order. (3) Level 3 (thematic engagement): Students' research proposal discussions were coded for thematic focus on "future learning competencies," including key competencies such as communication and discussion skills, inquiry skills, and teamwork/collaboration. (4) Level 4 (method engagement): Given the course's emphasis on educational research methods, students' discourse was coded for content knowledge, focusing on elements such as variable definition and operationalization, literature review and theoretical framework, and research design and methodological approach. (5) Level 5 (product engagement): The final research proposals, which were required to be submitted to KF, were assessed for quality based on educational research criteria (Aga, 2023).

In addition, data concerning students' academic achievements were gathered. Academic achievements were determined by final course grades on a 100-point scale (M = 73.73, SD = 8.44), which comprised the average scores of 10 regular class quizzes, all class assignments, and the final exam. Correlation and multiple regression analyses were conducted to examine how students' five-level KF performance indicators—online engagement (word count), cognitive engagement, content engagement (educational research methods), proposal engagement (future learning competencies), and outcome engagement (research proposal quality)—predicted final academic achievements.

Result

KB engagement on KF

Students exhibited high levels of engagement in KF discussions. This was reflected across multiple levels of participation and cognitive contribution as detailed below.

Level 1 (behavior engagement). At the foundational level, each student produced an average of 3,574.19 words (SD = 2,628.12) for KF discussion through three activities: (1) Note-contributing (M = 21.21; SD = 16.12)—This activity primarily involves creating discussion content; (2) Note-reading (M = 148.24; SD = 247.14)—This activity enables deeper understanding of the discussion content; (3) Note-modifying (M = 82.40;

SD = 229.20)—This activity enhances the quality of the discussion content. Then, at higher levels, we also coded keywords from the discussion content to identify “cognitive engagement”, “thematic engagement”, “method engagement”, and “proposal engagement”.

Level 2 (cognitive engagement). We assessed the level of students' cognitive engagement in KF discussion. Overall, higher-order engagement (analyzing, evaluating, and creating) accounted for 54.10% of related keyword contributions. In contrast, lower-order engagement (remembering, understanding, and applying) accounted for 45.90% of related keyword contributions. This finding suggests that KB practices fostered substantial engagement in deeper cognitive activities in KF discussions.

Level 3 (thematic engagement). We analyzed the level of discussion focus concerning the direction of students' research proposals. Specifically, we examined how discussions related to future learning competencies. The discussion showed that the most prominent dimensions were: (1) Communication and discussion skills (24.78%); (2) Inquiry skills (21.37%); and (3) Teamwork/collaboration (18.81%). This indicates that students' online proposal discussions regarding the research issues were highly relevant to future learning competencies, which represent an important research trend.

Level 4 (method engagement). We also examined the level of content richness in students' KF discussions in relation to "Educational Research Method"—the subject matter in this course. Students most frequently focused on three areas: (1) Variable definition and operationalization (24.56%); (2) Literature review and theoretical framework (17.97%); and (3) Research design and methodological approach (17.08%). These findings demonstrate their engagement in discussing the core knowledge required in the educational research methods course.

Level 5 (product engagement). At the final and highest level, we assessed the scores (M = 72.60; SD = 11.00) of students' final draft proposals. These proposals were uploaded to the KF platform and served as a culminating outcome of students' online KF discussion efforts.

Correlations between KF performance and academic achievements

Correlational analyses showed that academic achievements were significantly associated with the five engagement levels of online KF discussion. Regression analysis further demonstrated important findings. The combined 5-level factors/predictors accounted for 43.6% of the variance in academic achievements (with Adjusted $R^2 = .382$, $F(5, 52) = 8.04$, $p < .001$). Findings suggest that students' active engagement across multiple levels of KF discourse significantly predicted academic performance, accounting for 43.6% of the variance.

Discussions and conclusion

This study validates the effectiveness of Knowledge Building (KB) pedagogy in a higher education research methods course by focusing on actual online engagement and its impact on academic success. The results from the multi-level analysis of Knowledge Forum (KF) discourse demonstrate that students exhibited robust involvement, with 54.10% of contributions representing higher-order cognitive activities such as analyzing, evaluating, and creating. Furthermore, this multi-level engagement across behavioral, cognitive, thematic, methodological, and product dimensions significantly predicted 43.6% of the variance in final academic achievement.

Crucially, the findings underscore that the KF environment served as a vital arena for promoting Collaborative Problem Solving (CPS). By providing an asynchronous digital space for sustained "idea improvement," KF allowed students to manifest core CPS dimensions—such as social regulation, perspective-taking, and collective knowledge building—through their online interactions. The thematic analysis of discourse revealed that discussions were highly relevant to future learning competencies, particularly communication and teamwork skills, which are fundamental to effective CPS. This demonstrates that the KB environment on KF successfully transformed traditionally individualistic learning into a collaborative community responsible for collective knowledge advancement.

Practically, these results suggest that university educators should deliberately design online environments like KF to scaffold multiple levels of engagement, ensuring that collaborative tasks are aligned with both disciplinary mastery and 21st-century competencies. Although limited by its single-course scope, this study provides a validated framework for curriculum innovation in higher education. In conclusion, KB pedagogy, facilitated through the KF environment, offers an empirically grounded approach to fostering students' collaborative problem-solving (CPS) capabilities while simultaneously enhancing their academic excellence.

References

Aga, F. J. (2023). Determinants of academic research quality: The case at three Ethiopian public universities. *Social Sciences & Humanities Open, 7*(1), Article 100409.

Bereiter, C., & Scardamalia, M. (2014). Knowledge building and knowledge creation: One concept, two hills to climb. In S. C. Tan, H.-J. So, & J. Yeo (Eds.), *Knowledge creation in education* (pp. 35–52). Singapore: Springer.

Care, E., Scoular, C., & Griffin, P. (2016). Assessment of collaborative problem solving in education environments. Applied Measurement in Education, 29(4), 250–264.

Chen, B., & Hong, H.-Y. (2016). Schools as knowledge-building organizations: Thirty years of design research. *Educational Psychologist, 51*(2), 266–288.

Chen, L., Yoshimatsu, N., Goda, Y., Okubo, F., Taniguchi, Y., Oi, M., Konomi, S., Shimada, A., Ogata, H., & Yamada, M. (2019). Direction of collaborative problem solving-based STEM learning by learning analytics approach. *Research and Practice in Technology Enhanced Learning, 14*(1), Article 24.

Hesse, F., Care, E., Buder, J., Sassenberg, K., & Griffin, P. (2015). A framework for teachable collaborative problem solving skills. In P. Griffin & E. Care (Eds.), *Assessment and teaching of 21st century skills* (pp. 37–56). Springer.

Hong, H. Y., & Chiu, C. H. (2016). Understanding how students perceive the role of ideas for their knowledge work in a knowledge-building environment. *Australasian Journal of Educational Technology, 32*(1), 32–46.

Jiang, J.-P., Hu, J.-Y., Zhang, Y.-B., & Yin, X.-C. (2023). Fostering college students' critical thinking skills through peer assessment in the knowledge building community. *Interactive Learning Environments, 31*(10), 6480–6496.

Lin, K.-Y., Hong, H.-Y., & Chai, C. S. (2014). Development and validation of the knowledge-building environment scale. *Learning and Individual Differences, 30*, 124–132.

OECD. (2017). *PISA 2015 assessment and analytical framework: Science, reading, mathematic, financial literacy and collaborative problem solving* (Revised ed.). OECD Publishing.

Pásztor-Kovács, A., Pásztor, A., & Molnár, G. (2023). Measuring collaborative problem solving: Research agenda and assessment instrument. *Interactive Learning Environments, 31*(8), 5159–5179.

Scardamalia, M. (2002). Collective cognitive responsibility for the advancement of knowledge. In B. Smith (Ed.), *Liberal education in a knowledge society* (pp. 67–98). Open Court.

Scardamalia, M., & Bereiter, C. (2003). Knowledge building. In J. W. Guthrie (Ed.), *Encyclopedia of Education* (2nd ed., pp. 1370-1373). Macmillan Reference.

Scardamalia, M., & Bereiter, C. (2006). Knowledge building: Theory, pedagogy, and technology. In K. Sawyer (Ed.), *Cambridge Handbook of the Learning Sciences* (pp. 97-118). Cambridge University Press.

Scardamalia, M., & Bereiter, C. (2021). Knowledge building: Advancing the state of community knowledge. In U. Cress, C. Rosé, A. F. Wise, & J. Oshima (Eds.), *International handbook of computer-supported collaborative learning* (pp. 261-279). Springer.

Soliman, D., Costa, S., & Scardamalia, M. (2021). Knowledge building in online mode: Insights and reflections. *Education Sciences, 11*(8), Article 425.

Tan, S. C., Chan, C., Bielaczyc, K., Ma, L., Scardamalia, M., & Bereiter, C. (2021). Knowledge building: Aligning education with needs for knowledge creation in the digital age. *Educational Technology Research and Development, 69*(4), 2243–2266.

Tian, Q., & Zheng, X. (2024). Effectiveness of online collaborative problem-solving method on students' learning performance: A meta-analysis. *Journal of Computer Assisted Learning, 40*(1), 326–341.

Tong, Y., & Chan, C. K. K. (2023). Promoting knowledge building through meta-discourse and epistemic discourse understanding. *International Journal of Computer-Supported Collaborative Learning, 18*(3), 353–391.

Tsai, I. C., & Hsu, F. S. (2024). Effect of education based on creative problem-solving and asynchronous online discussions on student creativity in a product design course. *Thinking Skills and Creativity, 54*, Article 101650.

Zhang, J., Hong, H.-Y., Scardamalia, M., Teo, C. L., & Morley, E. A. (2011). Sustaining knowledge building as a principle-based innovation at an elementary school. *The Journal of the Learning Sciences, 20*(2), 262–307.

Zhang, S., Gao, Q., Sun, M., Cai, Z., Li, H., Tang, Y., & Liu, Q. (2022). Understanding student teachers' collaborative problem solving: Insights from an epistemic network analysis (ENA). *Computers & Education, 183*, Article 104485.

International Society of the Learning Sciences

Examining Socio-Cognitive and Affective Dynamics of Team Functioning and Leadership in Medical Simulations

Annamarie Brosnihan, University of Central Florida, annamarie.brosnihan@ucf.edu
Megan Wiedbusch, University of Central Florida, megan.wiedbusch@ucf.edu
Cameron Marano, University of Central Florida, cameron.marano@ucf.edu
Lindsey Schulz, Nemours Children's Hospital, lindsey.schulz@nemours.org
Anna Pirner, Nemours Children's Hospital, anna.pirner@nemours.org
Shiva Kalidindi, Nemours Children's Hospital, shiva.kalidindi@nemours.org
Roger Azevedo, University of Central Florida, roger.azevedo@ucf.edu

Abstract: Teamwork and leadership are critical in promoting team success and positive outcomes in high intensity contexts. Given the urgency during pediatric medical emergencies, it is crucial to understand what supports or impedes successful team dynamics, including cognitive, metacognitive, affective, motivational, and social (CAMMS) processes. Thus, this study explores the dynamics of first-, second-, and third-year medical residents' (N = 21) CAMMS processes and their influence on individual and team performance in three team-based pediatric emergency simulated scenarios. Results indicate a significant correlation between motivation and cognitive and social processes, as well as a significant influence of team performance on CAMMS dynamics. Our findings support the need to further examine the impact of individual level processes on medical team functionality and performance. Using multimodal data, future studies will continue to explore relationships involving CAMMS dynamics and medical team functioning and performance.

Introduction: Self- and socially shared regulated learning in medical teams

Medical simulation as a learning modality allows medical trainees to safely practice interprofessional and interpersonal skills, such as teamwork and leadership, in strenuous, high-stakes environments (Murphy et al., 2016). Efficient teamwork is pivotal to team success and positive outcomes, especially in high intensity contexts such as acute pediatric medical environments (Schmutz et al., 2019). To understand what facilitates or impedes efficient teamwork, it is essential to consider the (dis)engagement of various cognitive, affective, metacognitive, motivational, and social processes (CAMMS; Azevedo et al., 2018) of each team member and how they may either enhance or hinder individual and team performance (Giannakos et al., 2019). Despite the collaborative nature of healthcare teams, previous research in medical education has emphasized individual performance and SRL enactment; however, there remains a lack of understanding regarding how each team members' CAMMS processes play a role in their SSRL and the overall performance of healthcare teams (Bransen et al., 2021).

Socially shared regulated learning (SSRL) requires the collaborative activation and regulation of CAMMS processes between multiple individuals working together towards a common goal (Järvela et al., 2023). Effective SSRL is highly contextual (Järvenoja et al., 2015), requiring shared mental models (Kazemitabar et al., 2022) of both individual and joint task goals, defining expectations in accomplishing shared objectives and managing team dynamics (Hadwin et al., 2018). While SRL persists within the parameters of a designated team role, these efforts feed a collective objective, requiring the recognition and co-regulation of team-level behaviors (Hadwin et al., 2018; Bransen et al., 2021). As such, medical learners (i.e., students and residents) must carefully balance both their individual roles and expectations and their role as team members in the context of effective team management and collaborative learning (Bransen et al., 2021). Given the criticality of teamwork on team performance, there remains a need to investigate and distinguish the impact of each member's CAMMS processes on overall team functioning and performance (Schmutz et al., 2019). This study intends to identify the relevance of individual CAMMS processes and their influence on individual and team performance within team-based acute emergency medical simulations using low-fidelity manikins.

Methodology

Participants and simulated scenarios

 © ISLS

First-, second-, and third-year medical residents (N = 21; M_{age} = 31.5) specializing in pediatrics and emergency medicine were recruited from an American university teaching hospital to complete three simulated acute pediatric emergency scenarios (e.g., newborn with Meconium Aspiration Syndrome, Congenital Heart Defect, and Supraventricular Tachycardia) in a children's hospital simulation suite using low-fidelity medical manikins. Each simulated scenario was facilitated and evaluated by subject matter experts by providing standardized information about the simulated patient and necessary lab results when prompted.

Procedure

Once consent was provided, participants completed pre- self-report measures (see Table 1) and then were randomly placed into teams of four, comprising of three medical residents and one confederate nurse. Participants completed the three simulated scenarios with their respective teams, but alternated roles (team leader, primary assessment, and airway management) at the beginning of each scenario, assuming one role at a time.

The team leader in each simulated scenario was instrumented for collection with multimodal data (i.e., portable eye tracking glasses, electrodermal activity [EDA]) to capture fluctuations in CAMMS processes in real time and video/audio recordings captured entire team performances from multiple angles. Each scenario trial followed an identical paradigm: (1) roles were assigned and the team leader was instrumented; (2) the team entered the simulation room and instructed to stand in designated areas based on the role; (3) a pre-recorded verbal medical brief was presented to participants; (4) the simulation began and each team was given a maximum of seven minutes to complete each case. The scenario was considered complete when medical intervention was successful, indicated by the facilitators or seven minutes had elapsed. Each simulation room contained a low-fidelity manikin connected to monitors that accurately represented the patient's condition throughout the scenario. Following the completion of each simulation scenario, participants were given three minutes to decompress while the team leader was de-instrumented, new roles were assigned, and the new team leader was instrumented. This process was repeated until each participant assumed each role one time, and the three scenarios were simulated. Following the simulation of all three scenarios, participants completed post- self-report questionnaires (see Table 1), followed by a debrief led by medical experts.

Coding and scoring

Self-report measures were administered before and after the simulated scenarios (see Table 1). To evaluate changes in participant's CAMMS processes from pre- to post-simulation, difference scores were calculated and used in subsequent analyses.

Leadership performance was assessed using the Concise Assessment of Leadership Management instrument (CALM; Nadkarni et al., 2018). The CALM consists of 15 items rated on a 4-point Likert scale and evaluates four elements of leadership: (1) role/style; (2) communication; (3) team management; and (4) medical management. SMEs reviewed video recordings of each simulated scenario post-hoc and rated the team leader's performance, ranging from 16-60.

Team Performance was operationalized as the total duration of each simulated case. Simulated scenarios continued until medical intervention was successful, or seven minutes had elapsed.

Table 1
Description of Self-Report Measures

CAMMS Construct	Instrument	Reliability	Number of Items	Scale Size	Subscales
Affect	Medical Emotion Scale (MES; Duffy et al., 2020)	0.88 – 0.92	22 items	5-point Likert	Positive Activating Positive Deactivating Negative Activating Negative Deactivating Neutral
(Meta)Cognitive & Affect	Emotion Regulation Questionnaire (ERQ; Gross & John, 2003)	0.82 – 0.91	6 items	7-point Likert	*Cognitive Reappraisal
Cognitive and Social	Interprofessional Collaborative	0.95 – 0.97	20 items	5-point Likert	Collaboration Roles & Responsibilities

 © ISLS

	Competency Attainment Scale (ICCAS; MacDonald et al., 2010)				Collaborative Patient/Family-Centered Approach Conflict Management Resolution Team Functioning Collaboration
Motivation	Confidence Scale (C-Scale; Grundy, 1993)	0.93 – 0.94	5 items	5-point Likert	N/A

*While this is not the only subscale within the instrument, it is the only subscale administered in this study.

Results

RQ1: Do distinct relationships occur between changes in participant's self-reported CAMMS measures from pre-simulation to post-simulation?

A Pearson correlation matrix examining the relationships between changes in participant's self-reported CAMMS measures from pre-simulation to post-simulation indicated a strong significant negative relationship between changes in participant's reported confidence and interprofessional collaborative competencies (collaboration, $r(19) = -0.68$, $p < 0.01$; communication, $r(19) = -0.65$, $p < 0.01$; roles and responsibilities, $r(19) = -0.61$, $p < 0.01$; conflict management resolution, $r(19) = -0.53$, $p = 0.01$; and team functioning, $r(19) = -0.52$, $p = 0.01$) from pre-simulation to post-simulation.

RQ2: Do distinct clusters emerge based on changes in participant's self-reported CAMMS measures from pre-simulation to post-simulation, and do these clusters differ in performance (e.g., leadership and team)?

Hierarchical cluster analysis with Euclidean distance identified two distinguishable clusters. Cluster 1 ($N = 13$) exhibited a significant negative change in reported confidence (see Table 2a) and positive change in interprofessional collaborative competencies (see Table 2b) from pre-simulation to post-simulation. Cluster 2 ($N = 8$) exhibited a slight increase in reported confidence, and a smaller, predominately negative change in reported interprofessional collaborative competencies from pre-simulation to post-simulation.

Table 2a

Mean Differences in Affective and Motivational Measures from Pre- to Post.

	MES_PosAct	MES_PostDeact	MES_NegAct	MES_NegDeact	MES_Neutral	Confidence
Cluster 1	0	0.85	-0.02	0.13	-0.27	-4.31
Cluster 2	0.22	0.81	0.09	0.09	0.12	0.38

Table 2b

Mean Differences in (Meta)Cognitive, Social, and Affective Measures from Pre- to Post.

	ERQ_CogReap	IP_Comm	IP_Collab	IP_RR	IP_CPFCA	IP_CMR	IP_TF
Cluster 1	0.28	0.66	0.87	0.63	0.62	0.67	0.77
Cluster 2	-0.33	0.08	0.04	-0.09	-0.04	-0.29	-0.13

Mann-Whitney U tests indicated a significant difference in team performance between clusters, $X^2(1) = 4.41$, $p = 0.04$, where Cluster 2 ($M_{Duration} = 286$ secs) completed the simulated scenarios faster than Cluster 1 ($M_{duration} = 380$ secs). However, a significant difference in leadership performance between clusters was not exhibited, $X^2(1) = 0.01$, $p = 0.92$.

RQ3: How do the changes in self-reported CAMMS measures influence the structure of participant clusters?

Given the significant differences between clusters in team performance, we then examined how each of the variables contributed to the structure of participant clusters. To address the highly correlated variables defined in Research Question 1, we conducted a principal component analysis (PCA) on changes in confidence and

 © ISLS

interprofessional collaboration competencies from pre- to post-simulation. PCA revealed two components that accounted for 67% of the total variance (PC1: 49.93%; PC2: 17.43%).

Using the identified PC's and the remainder of the self-reports, hierarchical cluster analysis with Euclidean distance identified two distinguishable clusters. Cluster 1 (N = 11) exhibited larger, positive changes in affect (e.g., MES subscales; positive and negative activating and deactivating; see Table 3a) and smaller, negative changes in social, (meta)cognitive, and motivational components (e.g., PCs and cognitive reappraisal subscale from ERQ; see Table 3b). Cluster 2 (N = 10) exhibited smaller changes in affect (e.g., MES subscales; positive and negative activating and deactivating) and larger, positive changes in social, (meta)cognitive, and motivational components (e.g., PCs and cognitive reappraisal subscale from ERQ).

Table 3a
Mean Differences in Affective Measures from Pre- to Post

	MES_PosAct	MES_PosDeact	MES_NegAct	MES_NegDeact	MES_Neutral
Cluster 1	0.16	1.1	0.05	0.20	-0.05
Cluster 2	0	0.55	-0.01	0.03	-0.2

Table 3b
Mean Differences in (Meta)Cognitive, Social, and Motivational Measures

	ERQ_CogReap	PC1	PC2
Cluster 1	-0.24	-1.45	-0.05
Cluster 2	0.37	1.6	0.06

To evaluate the contribution of individual variables to cluster structure, we conducted a leave-one-variable-out silhouette analysis. The silhouette coefficient increased slightly following the removal of each variable, resulting in a negative drop score. As such, the removal of each variable individually did not influence the stability of the defined clusters. Similar results were demonstrated when conducting a leave-one-variable-out silhouette analysis on the defined hierarchical clusters in Research Question 2. While each variable independently may not be significantly influencing the cluster structure, their combined influence may be more significant.

Discussion

To explore CAMMS processes in the context of emergency medical simulations, we examined the relationships between changes in participant's CAMMS processes from pre- to post-simulation. Our results indicate a strong inverse relationship between participants reported confidence and teamwork. We also examined if distinct participant profiles could be defined based on changes in reported CAMMS processes from pre- to post-simulation, and if they differed in both individual and team performance. Cluster 1 exhibited a significant decrease in confidence and a slight increase in teamwork competencies, and Cluster 2 exhibited an increase in confidence and a predominate decrease in teamwork competencies. While there was no difference in leadership performance between clusters, Cluster 2 solved the acute medical simulation scenarios significantly faster than those in Cluster 1. We interpret these findings as a reflection of how team performance influenced participants' reported confidence and teamwork competences. Given the distinct participant profiles and their differences in team performance, we aimed to determine which CAMMS processes contributed most to cluster structure and distinction. However, no specific variable was greatly influencing the defined clusters. We interpret these results as the interaction between CAMMS processes is more influential on cluster distinction, rather than viewing them individually. Our findings further push the narrative of highlighting the interaction between individual CAMMS processes and how they contribute to the overall functioning and performance of a healthcare team.

Conclusion and future directions

The study of socio-cognitive and affective dynamics in team-based medical simulations is an area of research that significantly advances the intersection of healthcare education and learning sciences (Huang et al., 2025). By embedding research on medical teams within the CSCL tradition of regulated collaborative learning, we can bring theoretical depth to simulation-based education and push the boundaries of CSCL theory itself. Future research should further investigate the dynamic interplay between individual-level CAMMS processes and collective team dynamics in high-stakes medical simulations. Examining the contagiousness of each team leader's CAMMS processes on and in response to their team members and other contextual cues (e.g., medical monitors, manikins, nurse interventions) will be critical in understanding leadership and team dynamics.

 © ISLS

References

Azevedo, R., Taub, M., & Mudrick, N. V. (2018). Understanding and reasoning about real-time cognitive, affective, and metacognitive processes to foster self-regulation with advanced learning technologies. In D. H. Schunk & J. A. Greene (Eds.), *Handbook of Self-Regulation of Learning and Performance* (2nd ed., pp. 254-270). Routledge. https://doi.org/10.4324.9781315697048-17

Bransen, D., Govaerts, M. J. B., Panadero, E., Sluijsmans, D. M. A., & Driessen, E. W. (2021). Putting self-regulated learning into context: Integrating self-, co-, and socially shared regulation of learning. *Medical Education, 56*(1), 29-36. https://doi.org/10.1111/medu.14566

Duffy, M. C., Lajoie, S. P., Pekrun, R., & Lachapelle, K. (2020). Emotions in medical education: Examining the validity of the Medical Emotion Scale (MES) across authentic medical learning environments. *Learning and Instruction, 70*, 101150. https://doi.org/10.1016/j.learninstruc.2018.07.001

Giannakos, M. N., Sharma, K., Pappas, I. O., Kostakos, V., & Velloso, E. (2019). Multimodal data as a means to understand the learning experience. *International Journal of Information Management, 48*, 108-119. https://doi.org/10.1016/j.ijinfomgt.2019.02.003

Gross, J.J., & John, O.P. (2003). Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. *Journal of Personality and Social Psychology*, *85*, 348-362. https://doi.org/10.1037/0022-3514.85.2.348

Grundy, S. E. (1993). The confidence scale. *Nurse Educator, 18*, 6-9.

Huang, X., Nguyen, A., & Lajoie S. P. (2025). Examining socially shared regulation of learning in medical training: The interplay of heart rate change points on regulatory interactions. *European Journal of Psychology of Education, 40*(3), 93-118. https://doi.org/10.1007/s10212-025-01001-0

Järvenoja, H., Järvelä, S., & Malmberg, J. (2015). Understanding regulated learning in situative and contextual frameworks. *Educational Psychologist, 50*(3), 204-219. https://doi.org/10.1080/00461520.2015.1075400

Kazemitabar, M., Lajoie, S. P., & Doleck, T. (2022). A process model of team emotion regulation: An expansion of Gross' individual ER model. *Learning, Culture, and Social Interaction, 33*. https://doi.org/10.1016/j.lcsi.2022.100612

MacDonald, C., Archibald, D., Trumpower, D., Casimiro, L., Cragg, B., & Jelly, W. (2010). Designing and operationalizing a toolkit of bilingual interprofessional education assessment instruments. *Journal of Research in Interprofessional Practice and Education 1(*3). https://doi.org/10.22230/jripe.2010v1n3a36

Murphy, M., Curtis, K., McCloughen, A. (2016). What is the impact of multidisciplinary team simulation training on team performance and efficiency of patient care? An integrative review. *Australasian Emergency Nursing Journal 19*(1), 44-53. https://doi.org/10.1016/j.aenj.2015.10.001

Nadkarni, L. D., Roskind, C. G., Auerbach, M. A., Calhoun, A. W., Adler, M. D., & Kessler, D. O. (2018). The development and validation of a concise instrument for formative assessment of team leader performance during simulated pediatric resuscitations. *Simulation in Healthcare: Journal of the Society for Simulation in Healthcare 13*(2), 77-82. https://doi.org/10.1097/SIH.0000000000000267

Schmutz, J. B., Meier, L. L., & Manser, T. (2019). How effective is teamwork really? The relationship between teamwork and performance in healthcare teams: A systematic review and meta-analysis. *BMJ Open, 9*(9). https://doi.org/10.1136/bmjopen-2018-028280

International Society of the Learning Sciences

Leveraging Generative AI for Iteration in Undergraduate Design

Inés Álvarez-Icaza, Claudia Susana López Cruz, Martha Elena Núñez López, Griselda Esthela Oyervides Ramírez, Juan Carlos Márquez Cañizares, Christiam Ivan Mendoza García
i.alvarezicaza@tec.mx, lopezclau@tec.mx, martha.nunez@tec.mx, gris.oyervides@tec.mx, jcarquez@tec.mx, christiam@tec.mx
Tecnológico de Monterrey

Abstract: Training designers requires not only technical skills, but also the ability to think critically, creatively, and strategically. Design applications powered by Generative Artificial Intelligence (GenAI) are emerging every day, hence educators are concerned about the development of students' creative abilities. The project is a pedagogical experience mediated by a GenAI tool, the 'DIAgent'. Through a dynamic interaction between initial sketches, guiding questions, and automated feedback provided by the tool, students iterate their proposals, considering key variables such as form, usability, ergonomics, materials, and technology. This study is conducted with a mixed-method approach as an exploratory study to assess the creative abilities of a sample of 103 undergraduate students after the intervention. This paper presents the tool and the application protocol to be applied in six different design classes as the study's first phase results, leading to a better understanding of how to redefine the students' design process.

Introduction

Training designers demand more than technical proficiency; it requires cultivating critical, creative, and strategic thinking to address complex social, technological, and environmental transformations (Aldoseri et al., 2024). While there are some concerns regarding the use of emerging technologies as professional tools, younger designers might perceive them as essential. Therefore, the is an increasing demand for a redesigned framework of educational strategies mediated by technology. In this sense, it becomes key to integrating diverse reasoning modes and promoting high value-added innovation (Büler et al., 2022; Marrone et al., 2022).

This mixed-methods study examines a Generative Artificial Intelligence (GenAI)-mediated approach in design education, focusing on its role in enhancing student engagement and divergent thinking during ideation—an ability strongly supported in creativity research (De Bono, 2009; Runco et al., 2010; Palmiero et al., 2022; Saretzki et al., 2024). Through structured interaction with sketches and automated feedback, students iteratively refine proposals while applying deductive, inductive, and abductive reasoning (Kennedy & Thornberg, 2018). The study hypothesizes that GenAI strengthens conceptual development and competencies such as Reasoning for Complexity and Design Concept (Tecnologico de Monterrey, 2019). Ongoing validation evaluates whether AI improves both idea quantity and quality while fostering ethical technological integration.

Theoretical framework

Assessing creative potential has been a longstanding priority in education, recognized as a fundamental competence alongside critical thinking and collaboration (Thornhill-Miller et al., 2023; Zheng et al., 2025). Early frameworks developed by Torrance (1966) and Guilford (1967) laid the foundation for creativity assessment, particularly through the Torrance Test of Creative Thinking (TTCT), which measures Flexibility, Fluency, Originality, and Elaboration as essential cognitive dimensions (Almeida et al., 2008).

In the context of Generative Artificial Intelligence (GenAI), new concerns arise regarding the preservation of students' ability to produce original and meaningful outcomes (Chiodo, 2026), especially given qualities linked to authentic creativity such as consciousness, emotional depth, and lived experience (Da Pelo, 2025). In design education, GenAI is commonly approached as a tool to accelerate processes, generate AI-based products, or enhance creative and critical thinking. This study prioritizes the latter, emphasizing divergent thinking (De Bono, 2009; Álvarez, 2010; Wadinambiarachchi et al., 2024), creative fluency (Fernandez-Abascal & Martín Díaz, 2013), and requirement-based decision-making (Watkins et al., 2021). Although GenAI may strengthen TTCT-related factors (Habib et al., 2024), concerns about dependency persist. Balanced models such as co-creativity (McKayle, 2025) and AI-enhanced human abilities (Agboola & Yassin, 2025) advocate augmentation rather than replacement. In this sense, Kosmyna et al. (2025), observed and recognized cognitive load during an essay writing task with a small sample of participants, showing that "Brain-only participants" had more distributed networks, while Large

 © ISLS

Language Models (LLM) users displayed the weakest connectivity, which in the long term, translated to a poorer performance at neural, linguistic, and behavioral levels.

Method

This exploratory mixed-methods study (Creswell & Creswell, 2005) analyzed 103 design students at Tecnologico de Monterrey, Mexico, during the 2025 academic year. The objective was to implement the "DIAgent: Design with AI" tool to foster divergent thinking and deep iteration in sketching. Traditionally, teacher pressure to iterate generates tension in the studio. DIAgent acts as a guide in this critical process, allowing AI to direct the evolution of initial ideas while the professor focuses on the final evaluation. The intervention centered on the fifth semester (out of eight) but included groups from the fourth and sixth semesters, a control group, and a sample from another city to compare student disposition and the impact of context. This structure seeks to redefine the creative process, reduce interpersonal conflicts, and enhance the fluidity of proposals through generative technology.

Collecting initial data on participants

Each student developed a design concept from an initial sketch and iteratively refined it through guiding questions and automated feedback, addressing form, usability, ergonomics, materials, and technology. The experience aimed to strengthen critical thinking, continuous improvement, and informed decision-making. Assessment relied on the Creative Behavior Assessment Instrument (CBAI) (López Caudana et al., 2024), applied in week five to establish a baseline for four groups in this initial phase, with later stages expanding the sample and analysis. The CBAI, validated in Spanish within Mexican educational institutions, demonstrates high reliability (CVC = 0.971; Spearman-Brown reliability = 0.870). Results enable baseline comparison with post-intervention outcomes. Each student may obtain a maximum score of 100 points. For the evaluation of results in this application, the values for scoring on the Torrance test (1969) are considered, where:

- *High percentiles (≥ 75)* → High creative ability in the dimension evaluated.
- *Middle percentiles (25-74)* → Average creative development, with areas that can be improved.
- *Low percentiles (< 25)* → Need for creative stimulation interventions or strategies.

Measuring the impact of the intervention

The CBAI measures levels of creative behavior in students in the dimensions defined by Torrance (1966). These dimensions will be assessed in an evaluation conducted after the educational intervention, in where students' results will be scored based on their performance in the course and during the design process. Table 1 describes how scores are established for each dimension. To this point, the CBAI has been applied to almost half of the sample (N=43).

Table 1

Dimensions Addressed in the Study.

Dimension	Description	How it is scored
Fluency	Total number of relevant and meaningful ideas that a person produces in response to a stimulus.	One point is awarded for a minimum of 15 sketches and one more point if they exceed it, up to 25 sketches. From 26 to 35 sketches, 3 points, and 4 points if exceed 36.
Flexibility	Variety of categories or changes in focus present in the responses. Number of different types of ideas are generated.	Responses are grouped into categories; more categories = higher score, one point per category.
Originality	Answer again: How creative do you consider yourself to be?	Points are awarded for different ideas within each category, one point for each different idea.

To assess the impact that the evaluation had on each student, a questionnaire must be administered and completed by each participant. The questionnaire is divided into three sections: fluency, flexibility, and originality. It is administered through an online survey during the class following the exercise. The maximum total score is 10 points for the three sections (Fluency = 4 points, Flexibility = 3 points, and Originality = 3 points). It is noteworthy that this second questionnaire had not been applied yet, as the students had not finished their projects by the time this manuscript was produced.

 © ISLS

The tool

The DIAgent is an interface designed to guide students' interaction with a Large Language Model (LLM)-based tool. Its development began with a simulator built in ChatGPT, which faculty members tested to refine the pedagogical strategy and adjust the application protocol. Once the student journey and sequence were optimized to support the design process and incorporate generic requirements adaptable across courses, full tool construction started. Although specialized product-visualization software was considered, the team prioritized accessibility by using institutional resources, such as a standard Google account.

The experience promotes critical thinking and creative exploration in sketching. After producing 40 initial sketches, students use DIAgent to overcome creative blocks. They input project data, upload a sketch with contextual details, and connect to the Gemini-2.0-flash model. The system generates prompts addressing aesthetics, ergonomics, usability, and materials. Students refine their sketches and receive final evaluations (1–5 scale) per dimension. A downloadable PDF documents the process, transforming AI interaction into reflective academic dialogue.

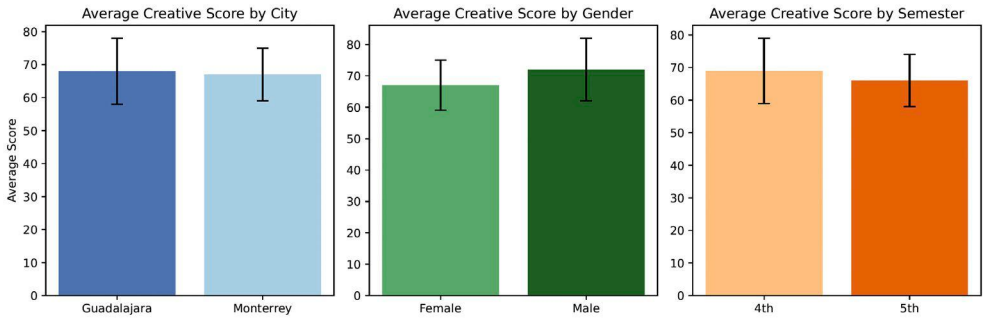
Results

Creative capacity score

Preliminary CBAI findings establish a quantitative baseline of creative performance and qualitative implementation insights. The partial sample (N = 43) included Industrial Design students from Monterrey (72%) and Guadalajara (28%), primarily fifth semester (74%) and fourth semester (26%). Gender distribution was 77% female, six male, and six undisclosed (excluded from gender analyses). The mean creative score was 67.68/100, indicating moderate creativity. One-way ANOVA revealed no significant differences by gender, campus, or semester. Although males showed slightly higher descriptive means, differences were not statistically significant, and semester level did not predict outcomes. Overall, the findings indicate a homogeneous creative baseline across demographic groups, supporting standardized intervention across campuses while prioritizing individualized creative development strategies over demographic distinctions. Figure 1 shows that males show slightly higher mean scores, but the difference is not statistically significant. Respondents stating that they prefer not to declare their gender (N=6) were eliminated from gender analysis. The figure also shows that both Monterrey and Guadalajara campuses display nearly identical creativity averages, and the semester of participants does not predict creative performance, as the differences are not significant.

Figure 1

Average Creative Scores.



Semantic trends

Qualitative analysis of four teachers' testimonials showed predominantly positive results (74%), with 18% neutral and 8% negative. Key findings indicate that students improved technical language for better tool interaction, engaged in critical decision-making through feedback, and found motivation by viewing the AI as a challenge to "beat."

Quantitative and sentiment findings indicate cultural readiness to integrate GenAI into creative development. Moderate creativity levels suggest opportunities to enhance divergent thinking. The study recommends variability-based feedback, adaptive pacing with micro-breaks, and structured prompts to reduce cognitive overload, supporting institutional adoption, and continuous technical improvement of the tool.

Table 2

Recurring Topics, Grouped by Sentiment and Cognition.

 © ISLS

Category	Description	Examples
A. Usability and Accessibility (highly positive)	Teachers emphasize the ease of use and the students' rapid adaptation	"easy to use," "students integrated it well," "there were not many questions"
B. Motivation and Gamification (positive and innovative)	Playful interaction with the AI promotes engagement and expressive improvement.	"they try to beat the artificial intelligence," "game strategy"
C. Technical and Reflective Development (constructive)	AI encourages precision, the use of technical language, and more reasoned redesign processes.	"use of technical language improves," "feedback on materials"
D. Workload and Saturation (neutral–negative)	Some teachers notice fatigue or repetitive comments during prolonged use.	"it became tiring," "repetitive comments"
E. Expectation of Autonomy and Reduced Friction (positive)	AI is perceived as a neutral mediator that decreases tensions in the learning process.	"no friction," "students receive these comments more openly"

Both datasets reveal that the institutional community is ready for these types of educational experiences, as students and teachers are emotionally receptive, yet creative output remains statistically even. Therefore, implementing the AI tool for divergent thinking should aim not just to develop a GenAI tool for design education, also to catalyse qualitative leaps in creative fluency, originality, and reflective reasoning.

Conclusions

As this study is ongoing, the analysis is not yet conclusive, although relevant insights have been gained regarding the impact of the Generative AI tool on the creative stage. Initial findings show no significant differences between student groups, establishing a consistent baseline for further research and formulating hypotheses without concern about the impact of context on the results. Several limitations are acknowledged: the sample size is small and not representative of the tool's full potential; furthermore, usability and technological acceptance have not been evaluated, nor is there evidence to determine whether the experience improved students' skills. Nevertheless, the qualitative analysis of the teachers' feedback is valuable in highlighting the importance of promoting deep reasoning in creative activities that consider aesthetic, ergonomic, and technical factors. Finally, the preliminary results underscore the need to expand the sample to more academic levels. Since the DIAgent tool does not record interactions, integrating learning analytics would be of benefit. Further evaluations are also required to strengthen the robustness of the changes in fluency, flexibility, and originality, considering the ethical and pedagogical implications of long-term use.

References

Agboola, O.P., Yassin, Y.N.H.M. (2025). AI Applications in Education: Enhancing Human Creativity Through Collaborative Design. In: Uden, L., Ting, IH. (eds) Knowledge Management in Organisations. KMO 2025. Communications in Computer and Information Science, vol 2563. Springer, Cham.

Aldoseri, A., Al-Khalifa, K. N., & Hamouda, A. M. (2024). AI-Powered Innovation in Digital Transformation: Key Pillars and Industry Impact. *Sustainability*, 16(5), 1790. https://doi.org/10.3390/su16051790

Almeida, L. S., Prieto, L. P., Ferrando, M., Oliveira, E., & Ferrándiz, C. (2008). Torrance Test of Creative Thinking: The question of its construct validity. *Thinking skills and creativity*, 3(1), 53-58.

Álvarez, E. (2010). *Creatividad y pensamiento divergente: Desafío de la mente o desafío del ambiente*. http://www.interac.es/adjuntos/crea_pensa_diver.pdf

Alvarez-Icaza, I., Suárez-Brito, P., Álvarez, J., & Molina-Espinosa, J. M. (2024). Relevance of objective and subjective profile: Creative behavior assessment in higher education students. *Frontiers in Education, 9*, Article 1387150. https://doi.org/10.3389/feduc.2024.1387150

Bühler, M. M., Jelinek, T., & Nübel, K. (2022). Training and Preparing Tomorrow's Workforce for the Fourth Industrial Revolution. *Education Sciences*, 12(11), 782. https://doi.org/10.3390/educsci12110782

Chiodo, S. (2026). Artificial Creativity?. *Minds & Machines*, 36, 6. https://doi.org/10.1007/s11023-025-09758-5

Creswell, J. W., & Creswell, J. D. (2005). Mixed methods research: Developments, debates, and dilemmas. *Research in organizations: Foundations and methods of inquiry*, 2, 315-326.

Da Pelo, M. (2025). Artificial creativity: can there be creativity without cognition?. *AI & Society: Knowledge, Culture and Communication*, https://doi.org/10.1007/s00146-025-02682-3

De Bono, E. (2009). *Lateral thinking: A textbook of creativity*. Penguin UK.

Fernandez-Abascal, E., & Martín Díaz, M. D. (2013). *Affective induction and creative thinking*. Creativity Research Journal, 25, 213–221. https://doi.org/10.1080/10400419.2013.783759

Guilford, J. P. (1967). Creativity: Yesterday, today and tomorrow. *The Journal of Creative Behavior*, 1(1), 3-14.

Guilford, J. P. (1973). *Characteristics of Creativity*. https://eric.ed.gov/?id=ED080171

Kennedy, B. L., & Thornberg, R. (2018). Deduction, induction, and abduction. *The SAGE handbook of qualitative data collection*, 49-64. http://digital.casalini.it/9781526416063

Kosmyna, N., Hauptmann, E., Yuan, Y. T., Situ, J., Liao, X.-H., Beresnitzky, A. V., Braunstein, I., & Maes, P. (2025, 10 de junio). Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for essay writing task. *MIT Media Lab*. https://www.media.mit.edu/publications/your-brain-on-chatgpt/

López Caudana, E. O., Ramírez Montoya, M. S., and Suárez Brito, P. (2024). Instrumento de Evaluación de Conducta Creativa (IECC) [Creative Behavior Assessment Instrument (CBAI)] (Record No. 03-2023060809274400-01). Tecnológico de Monterrey, Mexico.

Marrone, R., Taddeo, V., & Hill, G. (2022). Creativity and Artificial Intelligence—A Student Perspective. Journal of Intelligence, 10(3), 65. https://doi.org/10.3390/jintelligence10030065

McKayle, C.A. (2025). AI and Its Implications for Creativity Studies in Higher Education. In: Puccio, G.J., Lohiser, A., A. McKayle, C., Hertel, F. (eds) *The Palgrave Handbook of Applied Creativity in Global Higher Education*. Palgrave Studies in Creativity and Innovation in Organizations. Palgrave Macmillan, Cham. https://doi.org/10.1007/978-3-031-89816-7_35

Palmiero, M., Fusi, G., Crepaldi, M., Borsa, V. M., & Rusconi, M. L. (2022). Divergent thinking and the core executive functions: A state-of-the-art review. *Cognitive Processing*, *23*(3), 341–366.

Runco, M.A., Millar, G., Acar, S., & Cramond, B. (2010) Torrance Tests of Creative Thinking as Predictors of Personal and Public Achievement: A Fifty-Year Follow-Up, *Creativity Research Journal*, 22:4, 361-368. http://dx.doi.org/10.1080/10400419.2010.523393

Saretzki, J., Forthmann, B., & Benedek, M. (2024). A systematic quantitative review of divergent thinking assessments. *Psychology of Aesthetics, Creativity, and the Arts.* Advance online publication. https://doi.org/10.1037/aca0000691

Tecnologico de Monterrey (2019) Competencias transversales. Una visión desde el modelo educativo TEC21.

Thornhill-Miller, B., Camarda, A., Mercier, M., Burkhardt, J. M., Morisseau, T., Bourgeois-Bougrine, S., ... & Lubart, T. (2023). Creativity, critical thinking, communication, and collaboration: Assessment, certification, and promotion of 21st century skills for the future of work and education. *Journal of Intelligence*, 11(3), 54. https://doi.org/10.3390/jintelligence11030054

Torrance, E. P. (1969). Prediction of adult creative achievement among high school seniors. *Gifted Child Quarterly*, 13, 223 – 229. https://doi.org/10.1177/001698626901300401

Torrance, E. P. (1966). Torrance tests of creative thinking. Educational and psychological measurement. *APA PsycTests*. https://doi.org/10.1037/t05532-000

Wadinambiarachchi, S., Kelly, R. M., Pareek, S., Zhou, Q., & Velloso, E. (2024, May). The effects of generative ai on design fixation and divergent thinking. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (pp. 1-18). https://doi.org/10.1145/3613904.3642919

Watkins, M., Casamayor, J. L., Ramirez, M., Moreno, M., Faludi, J., & Pigosso, D. C. (2021). Sustainable product design education: current practice. *She Ji: The Journal of Design, Economics, and Innovation*, 7(4), 611637. https://doi.org/10.1016/j.sheji.2021.11.003

Wilson, R. N. (1954). Poetic Creativity†: Process and Personality. *Psychiatry*, 17(2), 163–176.

Zheng, W., Li, X. & Ben, S. (2025) Structural equation modelling of determining factors in musical creativity: an extended model based on Componential Theory of Creativity. BMC Psychol 13, 1090. https://doi.org/10.1186/s40359-025-03277-9

Acknowledgments

The authors would like to thank [BLINDED] for the financial support provided through the 'Challenge-Based Research Funding Program 2023', Project ID #IJXT070-23EG99001, entitled 'Complex Thinking Education for All (CTE4A): A Digital Hub and School for Lifelong Learners.'. Also, the support from the *Centro de Desarrollo Docente e Innovación Educativa* (CEDDIE) in the AISummit on summer of 2025, was key for the development of this project. Finally, the authors acknowledge the support of [BLINDED] colleagues of the School of Architecture, Art and Design, who were part of the conception of this project at the AISummit.

International Society of the Learning Sciences

From Regulation to Co-Regulation: A Game-Based Simulation for Human and Simulated Learners during Complex Problem Solving

Roger Azevedo, Megan Wiedbusch, Cameron Marano, Annamarie Brosnihan, Milouni Patel
roger.azevedo, megan.wiedbusch, cameron.marano, annamarie.brosnihan, milouni.patel@ucf.edu
University of Central Florida

Abstract: This paper presents *MetaSim*, a game-based simulation designed to foster the development and transfer of metacognitive and self-regulated learning (SRL) skills in human learners. Grounded in contemporary theories of SRL and metacognition, *MetaSim* targets algebra by integrating cognitive, affective, motivational, metacognitive, and social dimensions of learning and complex problem solving. Central to *MetaSim's* design, human learners interact with simulated learners (SLs), who serve as adaptive pedagogical agents (PA), embodying their human learner's SRL profile. While human learners engage in problem solving, the SLs model, explain, and justify SRL strategies, prompting learners to reflect on and refine their own regulatory behaviors. This paper presents a high-level overview of the four phases of teaching and testing the simulated learners' metacognitive capabilities in MetaSim.

Introduction

Developing robust SRL skills is crucial for student success in complex STEM fields that require strategic learning, problem-solving, reasoning, and adaptive expertise. SRL involves a cyclic process where learners plan, monitor, control, and reflect on their cognition, affect, metacognition, and motivation (CAMM) in response to internal conditions and task demands (Greene et al., 2024; Winne & Azevedo, 2022; Zimmerman & Moylan, 2009). Despite decades of research emphasizing the importance of SRL, many learners, particularly in STEM, lack the metacognitive skills necessary for sustained and effective regulation during problem-solving tasks (NASEM, 2025). Addressing this gap requires the development of learning environments, such as game-based AI simulations, that support SRL processes and assist learners in developing metacognitive and reflective skills, as well as adaptability for a sense of agency over how they learn (Azevedo & Wiedbusch, 2023). We address this issue by proposing the conceptualization of *MetaSim*, a new game-based AI simulation environment based on emerging research in the area of simulated learners (SLs; Käser & Alexandron, 2024; VanLehn et al., 1994). This paper presents a high-level overview of the four phases of teaching and testing the simulated learners' metacognitive capabilities in MetaSim.

Teaching and testing the SLs' metacognitive capabilities in MetaSim

MetaSim focuses on how human learners interact with, teach, observe, question, and learn *from* and *with* SLs using an AI-powered, game-based simulation environment, designed to foster metacognitive and SRL skills in the context of high school algebra problem solving (Azevedo et al., 2026). The overarching aim is to scaffold learners' own metacognitive growth by having them externalize and refine their knowledge of SRL processes through interactions with adaptive, intelligent SLs. In this section, we provide a brief description of how these interactions unfold across four dynamic, iterative, and cyclical phases.

Phase 1: Explicit teaching of SRL knowledge and strategies to simulated learners during algebra problem solving using multimodal trace data

In the first phase, learners take on the role of SRL instructors, teaching SLs specific metacognitive and SRL strategies (e.g., planning, goal-setting, monitoring, and self-evaluation) while solving algebra word problems. We will focus on complex algebra word problems, as they tend to be highly challenging for high school students (Borchers et al., 2025). These problems pose challenges for students as they require translating complex verbal information into mathematical and diagrammatic representations, applying appropriate cognitive strategies, solving them by engaging in multistep processes, and often inducing impasses or errors. Importantly, students must monitor each step for accuracy, which demands strong problem-solving and metacognitive skills, highlighting the need for their development. In addition, such problems require SRL skills to engage in complex problem solving by (1) understanding the problem, (2) using several effective cognitive strategies, (3) monitoring various aspects of the problem, (4) determining when to seek help, and (5) regulating one's emotions (e.g., confusion, frustration) and motivation (e.g., task value, task interest, self-efficacy) when faced with challenges such as impasses, incorrect or incomplete solutions, etc. (Borchers et al., 2025).

 © ISLS

During this phase, students teach their SL SRL strategies and demonstrate how to solve problems using these strategies. *MetaSim* collects and utilizes multimodal trace data (e.g., eye tracking, facial expressions, log files, clickstreams, dialogue interactions) to capture both the learner's teaching behaviors and the SLs' responses. Learners are prompted to verbalize and demonstrate declarative (*what*), procedural (*how*), and conditional (*when* and *why*) knowledge about SRL strategies, along with the algebra knowledge and skills needed to solve the problem. However, due to space limitations, this paper focuses only on SRL strategies. The multimodal data helps to create detailed profiles of the learner's instructional behaviors, supporting adaptive feedback loops between the learner and the SL.

This phase promotes the learner's ability to articulate and externalize their SRL knowledge and skills and fosters the SL's ability to acquire and encode these strategies. At the end of this explicit teaching phase, we expect the learner to have taught the SL everything they need to know to solve the algebra problem. Each student would be designing their own SL, and therefore, we envision a group of SLs, each reflecting and embodying the algebra and SRL skills of their corresponding human learners, populating *MetaSim* that will be used throughout the various phases described below.

In Phase 1, learners engage in a structured teaching mode where they instruct their SL on essential algebra problem-solving techniques alongside key SRL strategies. Core game mechanics include drag-and-drop scaffolding tools, content and strategy selection menus, and interactive dialogue trees, enabling learners to model and verbalize planning, monitoring, and evaluation processes. Leaners receive real-time multimodal feedback via gaze heatmaps, facial emotion recognition overlays, and gesture tracking, which visualize how SLs attend to and engage with the teaching process. A dynamic SRL dashboard tracks the use and sequencing of cognitive and metacognitive strategies, allowing learners to adjust their teaching and instruction based on the SL's evolving understanding and behavior. These mechanics encourage deliberate practice of SRL instruction, reinforcing how teaching others (especially the SLs) can enhance the human learner's own metacognitive awareness and instructional efficacy.

We acknowledge that learners will not always explicitly teach their SL everything they know and do not know. Additionally, the sequencing of algebra and SRL knowledge, processes, and skills that are taught may not be in an ideal sequence for optimal problem-solving. However, the human learners' multimodal data may reveal both (i) implicit SRL knowledge and skills that were not explicitly taught to the SL and (ii) a lack of specific SRL knowledge and skills. This data will be used in other phases to reveal, highlight, model, and scaffold learners' emerging SRL skills. Elaborations of data used for this and other phases described below are beyond the scope of this paper.

Phase 2: Learners observe SLs enact, model, and explain taught SRL strategies during problem solving

In the second phase, learners transition from teaching to observing. The SL now engages in algebra problem-solving by enacting the SRL strategies they were previously taught in phase 1 by the learner. These enactments are accompanied by real-time, metacognitive narration, where SLs explain what they are doing (e.g., "*I am identifying knowns and unknowns before solving*"), why they are doing it, and how it supports effective learning and problem-solving. Learners observe these behaviors through multimodal cues, including animated gestures, gaze shifts, and verbalized reflections. This observational learning context is designed to reinforce the effectiveness of strategies, promote metacognitive monitoring in the learner, and provide a comparative model of SRL in action. *MetaSim* may introduce challenges, distractors, or misconceptions that highlight when SRL strategies are applied appropriately or inappropriately; fostering learners' metacognitive monitoring and evaluation, and therefore prompt their metacognitive regulation strategies.

As each learner observes their SL enacting, modeling, and explaining the SRL strategies previously taught during algebra problem-solving. Through this observational (vicarious) learning phase, learners develop problem-solving skills such as identifying relevant information, selecting appropriate algebraic operations, and applying systematic procedures to solve equations (Dörner & Funke, 2017). Simultaneously, they strengthen SRL skills, including monitoring (e.g., detecting accurate SL strategy application, SL errors or strategy misapplication), evaluating (e.g., assessing the appropriateness of strategies used by SLs), and reflecting (e.g., comparing their own [i.e., taught] strategy use with that of the SL who may, in certain situations, fail to apply or misapply taught strategies). In addition, observing SLs verbalize their reasoning processes helps learners internalize, assess, and reinforce their *declarative knowledge* (i.e., a learner's explicit understanding of their own cognitive abilities, task demands, and available learning strategies), *procedural knowledge* (i.e., knowing how to implement them), and *conditional knowledge* (i.e., knowing when and why to use specific strategies) (Winne & Azevedo, 2022). This process fosters a deeper understanding of strategy transfer, metacognitive modeling, and the iterative nature of self-regulation. It ultimately enhances learners' ability to plan, regulate, and adapt their own learning in future

problem-solving tasks (Zimmerman & Moylan, 2009). It also emphasizes Winne's (2018) COPES (i.e., conditions, operations, products, evaluations, and standards) and SMART (i.e., search, monitor, assemble, rehearse, and translate) schemas for SRL. These models facilitate learners' development of internal standards and practice of cognitive evaluations for accurate self-regulation.

The instructional principles guiding this observational learning phase are grounded in social cognitive theory (Zimmerman & Moylan, 2009), cognitive apprenticeship (Collins & Kapur, 2022), and modeling-based instruction (Seel, 2003), emphasizing how learners acquire skills through guided observation, reflection, and internalization. First, the principle of modeling ensures that SLs explicitly demonstrate and verbalize SRL strategies such as planning, monitoring, and evaluating during algebra problem solving, allowing learners to witness effective cognitive and metacognitive behaviors in context. Second, scaffolded observation is used, where learners are supported by cues, prompts, and visualizations (e.g., strategy tags, feedback timelines) to focus attention on critical aspects of the SL's behavior. Third, reflection and explanation prompts encourage learners to engage in metacognitive reasoning by comparing the SL's actions to their knowledge and experiences. Fourth, the principle of strategy abstraction and generalization guides learners to move beyond specific examples to broader SRL principles they can apply across problems. Lastly, feedback-informed learning is employed, where learners receive targeted system feedback on their interpretations and reflections, reinforcing accurate understanding and promoting self-correction. Together, these principles foster the development of SRL and problem-solving skills by making otherwise internal strategies visible, explainable, and transferable through structured, guided observation.

During this phase, system features and game mechanics are designed to immerse learners in reflective and analytical thinking as they observe their SLs enact, model, and explain taught SRL strategies during algebra problem-solving. Key mechanics include interactive replay tools that enable learners to pause, rewind, and annotate the SL behaviors and highlighted SRL strategy tagging, which visually identifies when and how strategies such as planning, monitoring, and evaluation are used. An SRL strategy dashboard updates in real time to show which SRL strategies are being used, enabling learners to compare their own approaches with the SL's. Additionally, adaptive reflection questions guide learners to evaluate the effectiveness of the SL's decisions and consider how they might improve their own problem-solving and SRL approaches. These mechanics foster metacognitive awareness, strategy internalization, and skill transfer by turning observation into an active, scaffolded learning experience.

We expect this phase to take longer than Phase 1 and for learners to engage in many cycles of it, depending on several personal and educational goals ranging from instructional conditions, goals of teaching, learning and assessment, learners' SRL skills and expertise, and other individual differences. This cycle may also lead to learners revisiting Phase 1 as they teach their SL similar or more complex algebra word problems, without proceeding to Phases 3 and 4, described below.

Phase 3: Learners interrogate SLs about their SRL behaviors and prompt them to justify, refine, or improve strategy use

Phase 3 fosters metacognitive discourse between learners and SLs. Learners are encouraged to interrogate their SLs by asking them to justify their use of strategies, explain their rationale, reflect on their outcomes, and consider alternatives. For example, a learner might ask, "*Why did you decide to skip the step of checking your answer?*" or *Can you explain how you determined which strategy was most effective for this type of problem?*". Learners can also challenge their SL to try out more effective or nuanced strategies, pushing the SL to iterate and evolve. Multimodal trace data, including speech patterns, emotion detection, and response latency, are used to dynamically assess both the learner's questioning depth and the SL's metacognitive adaptability. This reciprocal teaching loop encourages learners to reflect on their own SRL repertoire while cultivating critical thinking, argumentation, and diagnostic reasoning. SLs will have "access" to detailed process models of SRL skills (e.g., Azevedo & Dever, 2022) that they can draw on to explain, justify, and model SRL skills; otherwise, the only SRL skills they will embody are those taught by their human learners. Due to space limitations, we are unable to describe this architectural feature in this paper.

As learners interrogate their SLs about their SRL behaviors and prompt them to justify, refine, or improve their strategy use, they acquire advanced metacognitive (e.g., monitoring use of strategies) and problem-solving skills that deepen their understanding of effective learning regulation. Through this active coaching role, learners develop strategic monitoring by identifying gaps or missteps in the SL's reasoning and evaluative judgment as they assess the appropriateness and timing of the SL's chosen strategies. This also strengthens the learner's conditional knowledge by enabling them to reason through *when* and *why* specific SRL strategies should be used, and how to adapt them based on task demands or performance outcomes. This process fosters reflective questioning, cognitive flexibility, and the ability to generate and test alternative strategies, as these are key

competencies for adaptive problem-solving (Spiro et al., 2019). By requiring learners to articulate feedback, challenge faulty assumptions, and co-construct improved solutions with their SLs, this phase promotes self-awareness, transfer of SRL knowledge, and executive control over learning processes, ultimately enabling learners to better regulate their own learning in future tasks.

The instructional principles guiding the interrogation phase are rooted in metacognitive scaffolding (Azevedo & Hadwin, 2005; Tabak & Reiser, 2022), Socratic questioning (duBoulay et al., 2023), and learning-by-teaching (Biswas et al., 2016) frameworks, designed to promote deep reflection, reasoning, and transfer of SRL strategies. Central to this phase is the principle of elaborative interrogation (Dunlosky et al., 2013), where learners are prompted to ask "why" and "how" questions that require the SL to justify and reflect on their strategy use, encouraging the learner to think critically about the underlying rationale of SRL behaviors. The principle of reciprocal teaching (Rosenshine & Maester, 1994) positions learners as cognitive coaches, reinforcing their own understanding by teaching and correcting others (i.e., SLs). Dialogic learning further supports this by fostering back-and-forth exchanges where learners evaluate explanations, identify weaknesses, and collaboratively refine strategies. Instruction is also guided by the principle of productive struggle (Kapur, 2024), where learners are challenged to analyze flawed reasoning or incomplete SRL strategies in SL responses, promoting the development of strategic thinking and error detection. Finally, guided self-explanation ensures that learners not only critique SL's choices but also articulate alternative strategies and justifications, strengthening their conditional knowledge and ability to generalize across tasks (Tan et al., 2025). These principles together create a high-engagement environment where learners internalize SRL processes by actively regulating another artificial learner's (i.e., their SL's) behavior.

In the interrogation phase of the game-based simulation, learners engage in an interactive dialogue with their SL, using branching question trees, voice or text-based input, and strategy-probing tools to ask SLs to explain, justify, or reconsider their use of SRL strategies during algebra problem solving. Game mechanics include dialogue-based scaffolding, where learners can choose from or generate prompts targeting specific SRL phases (e.g., *"Why did you choose that strategy?" or "Was that the best time to evaluate your progress?"*), and adaptive SL responses generated from a dynamic SRL reasoning model that reflects the SL's current understanding and learner profile. System features such as real-time strategy maps, confidence meters, and rationale logs support learners in evaluating the coherence and effectiveness of the SL's reasoning. A feedback engine guides learners to refine their questioning techniques by offering suggestions for deeper inquiry or highlighting missed opportunities to challenge flawed reasoning. This phase transforms learners into metacognitive coaches, reinforcing their own SRL skills by prompting critical thinking, strategy evaluation, and justification processes that deepen their understanding of when, why, and how to regulate learning.

Phase 4: SLs prepare human learners for more complex algebra problems

In the final phase, the roles shift again. The SLs now act as peers, coaches, or mentors, helping the human learners tackle more complex, multi-step algebra problems. These problems could involve advanced algebraic equations, word problems with multiple variables, or real-world applications of algebra. The SLs scaffold learners' planning, problem decomposition, and metacognitive monitoring in real time, drawing upon the strategies previously modeled and refined. They may prompt the learner to plan, reflect during the task, or self-evaluate their progress. The system uses real-time multimodal trace data to assess the learner's current SRL state and adapt the SL's guidance and feedback accordingly. This phase tests both the learner's internalization of SRL strategies and the SL's capacity to transfer and apply learned metacognitive behaviors to novel, more demanding contexts. By simulating adaptive peer support, the SLs serve as a metacognitive catalyst, encouraging learners to autonomously transfer and refine their SRL strategies in future learning situations.

This phase is significantly cognitively demanding, abstract, and strategic, and therefore focuses on qualitative shifts in learners' development of SRL skills. The time spent in this phase will differ based on individual student needs, instructional demands, teaching and instructional goals and objectives, and contextual factors. In addition, we envision a mode in this phase whereby learners activate other learners' SLs and run '*what if*' simulations, comparing their SL with those of other learners who embody more sophisticated SRL skills to ensure that they develop more task- and domain-specific independent SRL skills.

In summary, this four-phase process combines AI, game-based learning, and multimodal analytics to develop a rich, interactive learning environment. It not only improves the metacognitive skills of human learners but also assesses and enhances the SRL competencies of SLs, laying the foundation for a new era of intelligent, teachable agents that co-evolve with their human counterparts. SLs have the potential to revolutionize research into intelligent, AI-driven game-based simulations designed to prepare learners for complex workforce challenges and enhance their problem-solving skills.

 © ISLS

References

Azevedo, R., & Dever, D. (2022). Metacognition in multimedia learning. In R. E. Mayer & L. Fiorella (Eds.), *Cambridge handbook of multimedia* (3rd ed., pp. 132-141). Cambridge, MA: Cambridge University Press.

Azevedo, R., & Hadwin, A. F. (2005). Scaffolding self-regulated learning and metacognition: Implications for the design of computer-based scaffolds. *Instructional Science, 33*, 367–379.

Azevedo, R., & Wiedbusch, M. (2023). Theories of metacognition and pedagogy applied in AIED systems. In B. du Boulay, A. Mitrovic, & K. Yacef (Ed.), *Handbook of Artificial Intelligence in Education* (pp. 45-67). Edward Elgar.

Azevedo, R., Wiedbusch, M., Marano, C., Karimi, M., Brosnihan, A., Patel, M., Belga, J., & LaViola, J. (2026). *Designing simulated learners that externalize metacognition and self-regulation for complex problem solving*. Paper to be presented at the 28th International Conference on Human-Computer Interaction (HCII 2026), Montréal, Quebéc, Canada.

Biswas, G., Segedy, J. R., & Bunchongchit, K. (2016). From design to implementation to practice a learning by teaching system: Betty's Brain. *International Journal of Artificial Intelligence in Education, 26*, 350-364.

Borchers, C., Nagashima, T., Liu, P., Alibali, M. W., & Aleven, V. (2025). Is more gamification better? Evaluating playful interactions and narratives for algebra learning. In *Proceedings of the 20th European Conference on Technology Enhanced Learning (EC-TEL)*, Newcastle and Durham, UK.

Collins, A., & Kapur, M. (2022). Cognitive apprenticeship. In K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (3rd ed. pp. 156-174). Cambridge University Press.

Dörner, D., & Funke, J. (2017). Complex problem solving: What it is and what it is not. *Frontiers in Psychology, 8*,1153. doi: 10.3389/fpsyg.2017.01153

du Boulay, D., Mitrovic, A., & Yacef, K. (Eds.) (2023). *Handbook of Artificial intelligence in Education*. Edward Elgar.

Dunlosky, J., Rawson, K., Marsh, E., Nathan, M., & Willingham, D. (2013). Improving students' learning with effective learning techniques promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14*, 4-58.

Greene, J. A., Bernacki, M. L., & Hadwin, A. F. (2024). Self-regulation. *Handbook of educational psychology*, 314-334. Washington, DC: APA.

Kapur, M. (2024). *Productive failure: Unlocking deeper learning through the science of failing.* Wiley.

Käser, T. & Alexandron, G. (2024). SLs in educational technology: A systematic literature review and a Turing-like test. *International Journal of Artificial Intelligence in Education, 34*(2), 545-585.

National Academies of Sciences, Engineering, and Medicine. (2025). *Transforming undergraduate STEM education: Supporting equitable and effective teaching*. Washington, DC: The National Academies Press. https://doi.org/10.17226/28268.

Rosenshine, B., & Meister, C. (1994). Reciprocal teaching: A review of the research. *Review of Educational Research, 64*, 479-530.

Spiro, R., Feltovich, P. Gaunt, A., et al (2019). Cognitive flexibility theory and the accelerated development of adaptive readiness and adaptive response to novelty. In P. Ward, M. Schraagen, J. Gore, & E. Roth (Eds.), *Oxford handbook of expertise* (pp. 951-976). Oxford Press.

Tan, L., Gong, S., Wang, Y., Guo, X., Xu, X., & Wang, Y. (2025). Enhancing academic performance through self-explanation in digital learning environments (DLEs): A three-level meta-analysis. *Educational Psychology Review, 37* https://doi.org/10.1007/s10648-025-10001-x

Vanlehn, K., Ohlsson, S., & Nason, R. (1994). Applications of simulated students: An exploration. *International Journal of Artificial Intelligence in Education, 5*(2), 135–175.

Winne, P. H. (2018). Cognition and metacognition within self-regulated learning. In D. Schunk & Greene, J.A (Eds.), *Handbook of self-regulation of learning and performance* (2nd ed., pp. 254-270). New York, NY: Routledge.

Winne, P., & Azevedo, R. (2022). Metacognition and self-regulated learning. *The Cambridge handbook of the learning sciences, 3*, 93-113.

Zimmerman, B. J., & Moylan, A. R. (2009). Self-regulation: Where metacognition and motivation intersect. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 299–315). Routledge/Taylor & Francis Group.

Acknowledgments

This research was supported by funding from the National Science Foundation (DRL #2506427).

International Society of the Learning Sciences

Knowledge Network for Innovations in Learning and Teaching (KNILT): Teacher Learning Through Knowledge Creation

Ying Kong, Jianwei Zhang, Aya Isaac
ykong2@albany.edu, jzhang1@albany.edu, aisaac@albany.edu
University at Albany, State University at New York, Albany, NY, USA

Abstract: This research engages teachers as knowledge builders who create minicourses on the Knowledge Network for Innovations in Learning and Teaching (KNILT; https://knilt.arcc.albany.edu/) to support peer teacher learning around authentic issues of educational practice. A content analysis of 42 teacher-designed minicourses examined the types of knowledge addressed and the teaching methods employed to support teacher learning. The findings show that the minicourses devote substantial effort to helping teachers deepen their knowledge of teaching, student learning, curriculum content, and technology, with rich co-occurrence among knowledge themes and learning activities. KNILT operationalizes a learning-as-creating model for professional development, positioning teachers as knowledge builders who advance both their own and their peers' pedagogical expertise through iterative minicourse design and refinement.

Introduction

To prepare students for a rapidly changing world, education needs to develop new classroom practices that engage students in creative knowledge work as a collaborative community (Chinn & Duncan, 2021; Paavola et al., 2004). To develop such practices, teachers themselves must have opportunities to engage in creative knowledge work with colleagues so that they can embrace an identity as professional knowledge builders (Chan, 2011; Teo, 2023; Zhang et al., 2011; Vokatis & Zhang, 2016). Unfortunately, many current teacher professional development programs prioritize content delivery by external experts, which tends to limit teachers' creative input and has shown limited impact on instructional practice and professional growth (Ball & Cohen, 1999; Borko, 2004; Darling-Hammond & Richardson, 2009; Wilson & Berne, 1999; Yoon et al. 2007). The field thus needs to explore new ways to create participatory teacher learning environments that encourage teachers' agency for addressing authentic challenges and working on innovative learning designs drawing upon personal experiences and distributed expertise (Günal-Aggül et al., 2024).

The research project reported in this paper aims to create an open and participatory environment that engages teachers as pedagogical knowledge builders. Teachers create knowledge artifacts for peer teacher learning, focusing on authentic issues in educational practice. Building on research on professional knowledge-building communities, our design supports teachers in co-designing innovative practices and pursuing continual improvement through data-informed reflection (Chan, 2011; Ma & Scardamalia, 2022; Popp & Goldman, 2016; Teo, 2023; Zhang et al., 2011). As guiding principles, our design foregrounds *teacher agency* for addressing authentic problems/needs through knowledge-building efforts. They share their knowledge and insights through designing *public knowledge artifacts* for peer teacher learning. The knowledge artifacts are structured as minicourses, each with 3-5 learning units moving from pedagogical principles (know-what, know-why) to practice (know-how in classroom context), using design as a ladder for principle-based practical problem-solving in education (Bereiter, 2014). A participatory, open-access environment is designed to support teacher contribution and peer learning: Knowledge Network for Innovations in Learning and Teaching (KNILT, https://knilt.arcc.albany.edu)

KNILT is a long-term design-based research project implemented in an online teacher education program (MS-CDIT) at the University at Albany. The teacher learners participating in KNILT carry diverse expertise as in-service teachers, college instructors, technologists and designers, librarians, and other professionals. Over 700 minicourses have been created to support teacher learning in high-need areas related to pedagogical innovation, curriculum change, new assessments, and technologies. Initial designs of the minicourses were reviewed by a small group of peer teachers for feedback. After completion, the minicourses are shared on KNILT and continually used for peer teacher learning in the MS-CDIT program. The minicourses are further made freely available to educators from all around the world.

In the context of the KNILT, the current study intends to investigate how teachers build public knowledge in the form of minicourses for peer teacher learning. Our research questions ask: What kinds of teacher knowledge is addressed in their minicourses, using what teacher learning activities? Related to principle-based

 © ISLS

practical problem-solving among teachers, our analysis looks at how the minicourses connect pedagogical principles with classroom practice through developing principled practical knowledge (PPK): coherent principle-based accounts that offer practical guidance for classroom teaching (Bereiter, 2014).

Method

Dataset

From KNILT's more than 700 minicourses, we drew a purposive sample (Creswell & Poth, 2018) of 42 minicourses created in the five most recent semesters. Selection criteria focused on K–12 contexts and ensured coverage of multiple learning areas, including STEM, literacy and language, social-emotional learning, new technologies, and special education.

Data analysis

The main data source was the unit pages of the minicourses, supported by the design process documents created by the minicourse authors. We conducted systematic content analysis at the learning unit/module level. Drawing on trathe related literature (Bereiter, 2014; Darling-Hammond et al., 2024; Mishra & Koehler, 2006), the coding scheme addressed three dimensions: (1) areas of teacher knowledge—seven level 1 categories (Knowledge of Learning, Learner, Teaching, Teacher, Content Knowledge, Technology Knowledge, and Context/World Knowledge) with 168 level 2 subcategories; (2) types of teacher learning activities—four level 1 categories (Sources, Engagement, Reflection, Assessment) with 19 subcodes; and (3) theory-practice connections (principled knowledge, practical knowledge, and integrated PPK).

During the codebook development, two coders first independently coded 15 minicourses through open coding, followed by three iterations of discussion and consolidation. Formal intercoder reliability on 10 minicourses yielded Cohen's $\kappa = 0.82$ for knowledge (range 0.76–0.89) and $\kappa = 0.85$ for activities (range 0.79–0.91), indicating substantial-to-near-perfect agreement (Landis & Koch, 1977). Ambiguous segments in the remaining corpus (n = 23, 2.6%) were resolved through coder discussion. Co-occurrence matrices and PPK distribution visualizations were generated in MAXQDA (VERBI Software, 2026) to examine how different types of knowledge and activities are integrated in each minicourse.

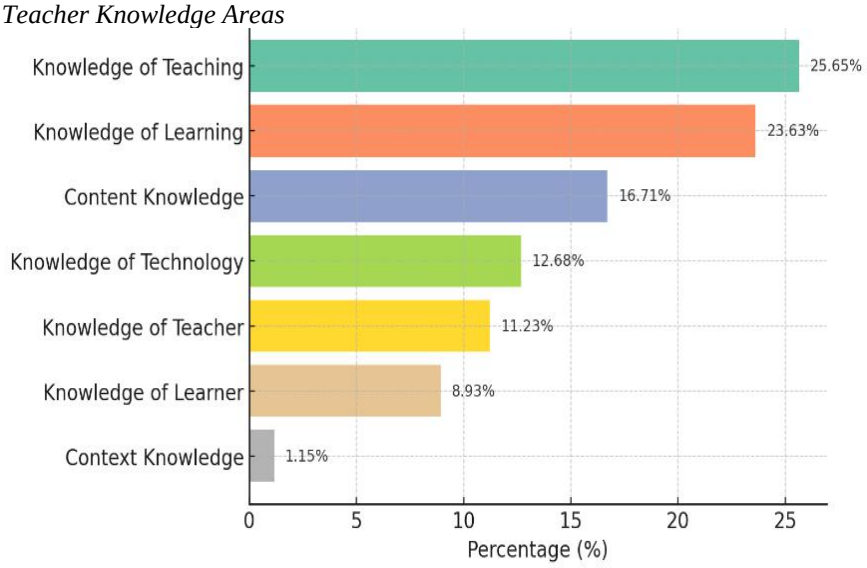
Findings

What kinds of teacher knowledge are addressed?

The content analysis applied a total of 886 knowledge-related codes to the 42 minicourses. Figure 1 shows the distribution of different areas of knowledge needs addressed by the minicourse units.

Figure 1

Teacher Knowledge Areas



 © ISLS

Among the level 1 categories, knowledge of teaching (25.65%) and knowledge of student learning (23.63%) are most extensively addressed, followed by content knowledge in changing curriculum areas (16.71%), and knowledge about technology use in classrooms (12.68%). A small proportion of the minicourses specifically focus on knowledge about teacher's professional and personal growth (11.23%) and knowledge about diverse learner populations (8.93%) and learning contexts that supports adaptive teaching. At the subcode level, designers most frequently address subject knowledge (7.61%), standards alignment (6.93%) in the area of content knowledge; teaching approaches (6.23%), principles of learning design (6.22%), and evaluation/assessment (6.24Fr%) under knowledge of teaching.

Theory-practice connections across minicourses

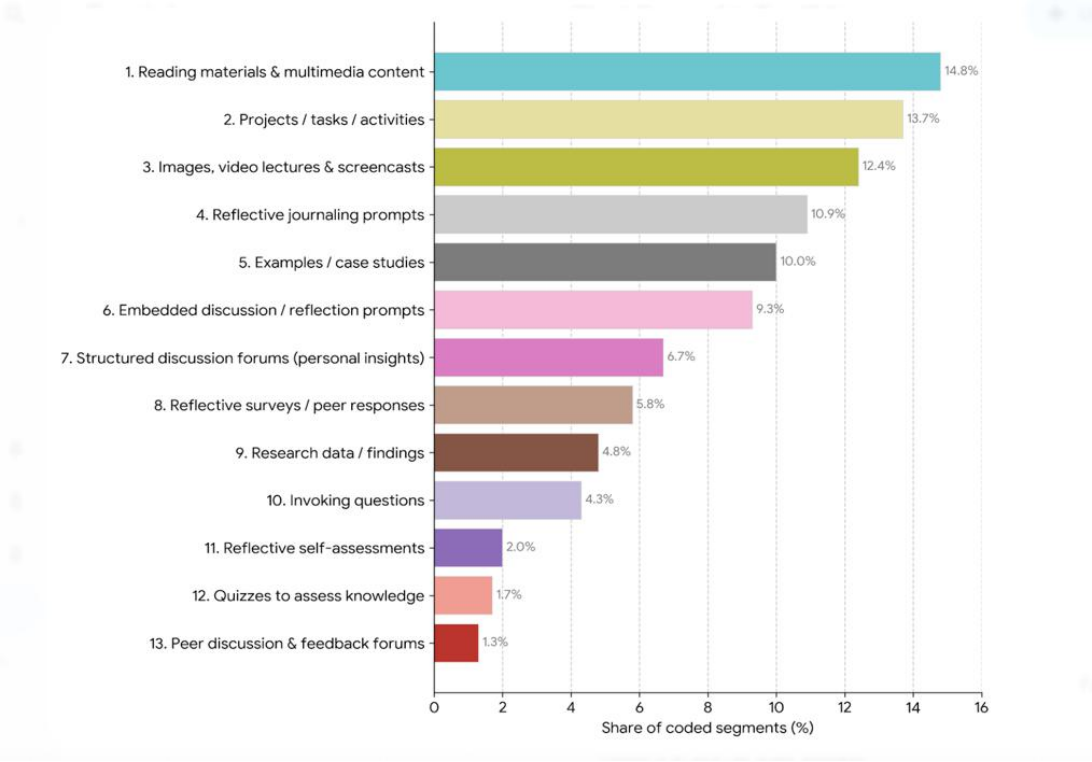
Principle knowledge (guiding theories and principles) appears in 100% of minicourses, practical knowledge in 93.3%, and PPK in 80%. Most minicourses follow a principle → practical → PPK sequence (86.7%), moving from principle-based knowledge to practical strategies in contexts to creating principle-based designs for practical problem solving. For instance, the minicourse *Utilizing Graphic Organizers* illustrates the deepening process: Unit 1 builds principled knowledge (what a graphic organizer is and why it works), Unit 2 addresses practical knowledge (how to create one with digital tools), and Unit 3 culminates in PPK—applying graphic organizers across grade levels in authentic classrooms, with the explanatory accounts presented in Units 1-2 guiding the design work in Unit 3.

What kinds of learning activities are integrated?

Figure 2 shows the diverse teacher learning activities used in the minicourses. The most frequent activities focus on purposeful reading of knowledge sources (42.12%, n=227), which includes reading materials and multimedia content, visuals and videos, research findings, and so forth. The next major category of activities supports interactive engagement (30.06%, n=162), such as project-based activities, reflective journaling, discussion, and survey. Rich teacher reflection activities are used (23.38%, n=126). As a limitation, formal assessment is comparatively rare (4.45%, n=24) partly due to the limitation of the KNILT technology platform. Sequential analysis of the sub-codes in each minicourse reveals a pronounced read-watch-make rhythm, building on reading materials/multimedia and video cases to work on authentic classroom design projects and challenges.

Figure 2

Frequencies of Teacher Learning Activities



 © ISLS

Co-occurrence of coding themes

Our analysis further examines how the different codes of teacher knowledge and learning activities co-occur within each minicourse. Codes of teacher knowledge cluster around pedagogical reasoning (e.g., Instructional Strategies appearing with Scaffolding) and learning activity codes bundle together to support flows of learning: Multimedia and Examples/Case Studies co-occur 18 times; Reading Materials and Examples/Case Studies 12 times; Project/Tasks with Invoking Questions 10 times. The combined use of different learning activities represents a strong strategy to support teacher sensemaking and designing.

Discussion

To advance designs of teacher learning through knowledge creation, this study examined 42 minicourses created by teachers on KNILT for peer teacher learning. In line with KNILT's guiding principles, the findings show that the teacher authors were able to create public-facing minicourses to address a wide range of knowledge needs among teachers (Figure 1), such as those highlighted in the literature (Bereiter, 2014; Darling-Hammond et al., 2024; Gavrilas & Kotsis, 2026). The minicourses help teachers deepen their knowledge of teaching, learning, disciplinary/interdisciplinary contents, and new technologies as related to their work with different student populations. A less addressed area of knowledge is to better understand the local and global contexts of teaching. Since the minicourse designers are themselves teacher learners and practitioners, their field-based experience helps them to understand their peers' needs and create accessible knowledge resources and learning activities. A rich set of teacher learning activities are involved to help peer teachers learn with the minicourses (Figure 2). The co-occurrence analysis reveals how the different activities are used in combination. The read-watch-make rhythm provides a learning flow in which teachers digest new knowledge through close reading, analyze lesson examples and classroom cases, and then work on authentic design to improve their classroom practices. In response to the persistent chasm between theory and practice in teacher education, the teacher-designed minicourses demonstrate productive patterns for integrating pedagogical principles with classroom practice. The flow of learning units in most of the minicourses allows teacher learners to build on principle-based knowledge to learn new practical strategies in specific contexts and eventually create principle-based designs they may adopt to improve their future teaching. Teacher designers with graduate-level education training seem well-positioned to synthesize different layers of knowledge and create peer learning materials that support PPK development, which is essential to adaptive teaching and continual improvement (Bereiter, 2014).

As limitations, the data analysis only provides a general view of the content and activity features in a small sample of the KNILT minicourses. Our ongoing analysis will further expand the analysis and examine teachers' professional design and knowledge-building efforts that lead to the minicourse creation. Future upgrading of the KNILT environment will also integrate tools to better support teachers' instructional design, including using generative AI as design partners to analyze authentic needs, synthesize the field's best knowledge and practices, and create engaging learning activities that support teachers' pedagogical sensemaking and design.

References

Ball, D. L., & Cohen, D. K. (1999). Developing practice, developing practitioners: Toward a practice-based theory of professional education. In L. Darling-Hammond & G. Sykes (Eds.), *Teaching as the learning profession: Handbook of policy and practice* (pp. 3–32). Jossey-Bass.

Bereiter, C. (2014). Principled practical knowledge: Not a bridge but a ladder. *Journal of the Learning Sciences*, 23(1), 4–17. https://doi.org/10.1080/10508406.2013.812533.

Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher*, 33(8), 3–15. https://doi.org/10.3102/0013189X033008003

Chan, C. K. K. (2011). Bridging research and practice: Implementing and sustaining knowledge building in Hong Kong classrooms. *International Journal of Computer-Supported Collaborative Learning, 6,* 147-186.

Chinn, C.A. & Duncan, R.G. (2021). Inquiry and Learning. In R.G. Duncan & C.A. Chinn (eds.*) International Handbook of Inquiry and Learning* (pp. 1–14). New York, NY: Routledge.

Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches (4th ed.).* Sage Publications.

Darling-Hammond, L., & Richardson, N. (2009). *Research review/Teacher learning: What matters? Educational Leadership*, 66(5), 46–53.

Darling-Hammond, L., Schachner, A. C. W., Wojcikiewicz, S. K., & Flook, L. (2024). Educating teachers to enact the science of learning and development. *Applied Developmental Science*, 28(1), 1–21. https://doi.org/10.1080/10888691.2022.2130506

Gavrilas, L., & Kotsis, K. T. (2026). A theoretical framework for understanding the effective educator: Defining characteristics, practices, and implications. *International Journal of Professional Development, Learners and Learning*, 8(1), e2602.

Günal-Aggül, Y., Horn, I., Shapiro, B. R., Öner, D., Metts, E., Chen, G. A., Shim, J., Esboldt, J., Harmon, M., Yoon, S. A., Aldemir, T., & Jurow, A. S. (2024). Designing Participatory Teacher Learning Environments: Inquiries, Methods, and Perspectives from Learning Sciences. In Lindgren, R., Asino, T. I., Kyza, E. A., Looi, C. K., Keifert, D. T., & Suárez, E. (Eds.), *Proceedings of the 18th International Conference of the Learning Sciences - ICLS 2024* (pp. 1910-1917). International Society of the Learning Sciences.

Ma, L., & Scardamalia, M. (2022). Teachers as designers in knowledge building innovation networks. *In The Learning Sciences in Conversation* (pp. 107-120). Routledge.

Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers college record*, *108*(6), 1017-1054.

Paavola, S., & Hakkarainen, K. (2005).The knowledge creation metaphor--An emergent epistemological approach to learning. *Science & Education*, 14(6), 535–557. https://doi.org/10.1007/s11191-004-5157-0

Paavola, S., Lipponen, L., & Hakkarainen, K. (2004). Models of innovative knowledge communities and three metaphors of learning. Review of Educational Research, 74(4), 557–576. https://doi.org/10.3102/00346543074004557

Popp, J. S., & Goldman, S. R. (2016). Knowledge building in teacher professional learning communities: Focus of meeting matters. *Teaching and Teacher Education*, 59, 347–359. https://doi.org/10.1016/j.tate.2016.06.007

Teo, C. L. (2023). *A teacher's playkit for knowledge building: Creative Ways to Make Students' Ideas Come Alive.* Routledge.

Vokatis, B., & Zhang, J. (2016). The professional identity of three innovative teachers engaging in sustained knowledge building using technology. *Frontline Learning Research*, 4(1), 58–77. https://doi.org/10.14786/flr.v4i1.223

Wilson, S. M., & Berne, J. (1999). Teacher learning and the acquisition of professional knowledge: An examination of research on contemporary professional development. *Review of Research in Education*, 24, 173–212. https://doi.org/10.2307/1167270

Yoon, K. S., Duncan, T., Lee, S. W.-Y., Scarloss, B., & Shapley, K. (2007). *Reviewing the evidence on how teacher professional development affects student achievement* (Issues & Answers Report, REL 2007–No. 033). U.S. Department of Education, Institute of Education Sciences.

Zhang, J., Hong, H.-Y., Scardamalia, M., Teo, C., & Morley, E. (2011). Sustaining knowledge building as a principle-based innovation at an elementary school. *Journal of the Learning Sciences*, *20*(2), 262–307.

Acknowledgments

We thank the contributing educators for sharing their professional knowledge and practices on KNILT (https://knilt.arcc.albany.edu/) as an open educational resource.

International Society of the Learning Sciences

Learning Across Levels: Conceptualizing Transitional Space

Xuesong Cang, Sadhana Puntambekar
xcang@wisc.edu, spuntambekar@wisc.edu
University of Wisconsin, Madison

Abstract: This paper conceptualizes *Transitional Space* as a framework for understanding how learning moves between individual and collective levels in *CSCL*. While prior research has revealed how groups construct shared meaning, how these collective advances become personally meaningful for individual learners is less articulated in a systematic way. *Transitional Space* reframes this relationship by conceptualizing learning as a relational movement in which collective knowledge construction and individually understanding continuously influence one another. The framework comprises two complementary facets: *enacted Transitional Space* capturing the interactional moments when group constructed ideas surpasses/below some members' current understandings and *supported Transitional Space* highlighting the designable conditions that support these moments through tools, scaffolds, and structured activities. Two facets link the emergent and designed dimension of cross-level learning, emphasizing how *Transitional Space* makes visible and actionable the dynamic processes through which collective knowledge becomes individual sense-making, and how design supports these movements across levels.

Introduction

Collaborative learning brings numerous benefits, yet classroom enactment remains difficult. Teacher must adopt unfamiliar practices, students must invest effortful participation, and assessment mostly privileges individual products, creating tension with collective work (Hakkarainen, 2009). Design guidance to address these persisting challenges is likewise fragmented, though research offers insights about task sequencing and tools (Celepkolu et al., 2017). Such practical considerations also reveal deeper theoretical challenges. First, how students internalize knowledge, artifacts, or other resources that are constructed in collaboration (Lawrence & Valsiner, 1993). Second, whether, and in what ways, individuals advance their own understanding when they externalize thinking in relation to the group's activity (Scardamalia, & Bereiter, 2021). Third, what are sufficient support to mediate social interactions in ways that enhance learning (Zhang et al., 2018). These questions are intricately linked, requiring a coherent framework that guides the design of learning environments. This work conceptualizes the framework of *Transitional Space* to addresses this challenge by reviewing existing perspectives, identifying their limitations, advocating for a more integrated approach that foregrounds the interdependence of individual and collective level of learning, and discussing the implications for supporting learning environment design and in-school collaborative learning practices.

What is cross-level learning in CSCL?

Define cross-level learning and its empirical evidence

Collaborative learning requires learners to participate in activities with others to achieve a desired goal (Rogoff et al., 1995). Individuals make sense of ideas of others while bringing theirs (Scardamalia & Bereiter, 2021). Thus, they constantly shift between individual and collective level of thinking (Stahl, 2013), no matter in a sequence of deliberately design activities or just within one activity. For the former, individual students may need to participate in one activity that requires them to think alone first followed by a shared-out or group work (Celepkolu et al., 2017). This requires them to switch the planes of activities through the structure of activities. For the latter, individuals engage in temporal transitions between individual thinking and collective reasoning and negotiating (Damşa et al., 2019). Therefore, learning across levels is the process through which knowledge, practice, and understanding emerge and transform through continuous transitions between individual and collective levels. Through participation in shared activity, personal recourses are externalized and reconstructed in social plane (Rogoff et al, 1995). Simultaneously, collective discourse serves as mediating structures through which individuals internalize shared meaning and extend their personal understanding (Vygotsky, 1978; Wertsch, 1991). As such, collaborative learning is neither exclusively internal nor purely social. It is co-constituted through dynamic coupling of individual sense-making and interactions at the collective level.

A substantial body of CSCL research also provides empirical evidence that learning unfolds across level, though often without explicit theorization of how learning traverse levels. Knowledge Building grounded studies

 © ISLS

unpack how knowledge is constructed collectively through individuals' continuous contribution. For example, Zhang et el., (2018) reveals that community knowledge influenced both their personal understanding and the collective advancement of ideas. Group cognition perspective illustrates that the shared meaning emerges through the coordination of ideas in discourse. Through the construction of shared problem spaces, collective reasoning is built upon the articulation and uptake of individual perspectives (Stahl, 2013, 2016). Research unpacks productive interactions show that effective collaboration depends on how individuals align contributions to sustain shared inquiry and joint problem solving. When coordination fails, even capable individuals may not benefit from group work (Barron, 2003). In addition, studies using Socially Shared Regulation and Group Awareness adds another layer to this evidence. Järvelä and Hadwin (2013) demonstrated that regulation process, including planning, monitoring, and adapting strategies, can occur collectively and support both group productivity and individual metacognitive development. Tools that visualize group progress foster awareness that connects individual engagement to collective outcomes (Schnauber & Bodemer, 2022). Together, these suggest that collaborative learning constantly requires students across levels through mediation of other, tools or the environment itself.

Why need to emphasize learning across-level?

Despite this empirical progress, we grapple with theoretical gaps that stem from the underlying assumptions of these studies, which can be considered as major approaches. First, the individual-centered approach mainly investigated how individual learning occurs as a result of collaborative learning. These studies usually root in cognitive and individual constructivist traditions, assuming that learning is the mental restructuring within individual (Chi, 2009). Therefore, individual is usually the unit of analysis with the assumption that collaboration provides stimuli or create condition for conceptual change. Consequently, the transition between group participation and individual understanding is treated as an implicit byproduct without theorized. Second, the group-focused approach emphasizes that the collectively constructed knowledge products/artifacts as the main indicator of learning. Studies usually ground into sociocultural, socio-material, socio-technical or distributed cognition perspectives (Pea, 1993; Vygotsky, 1978). They posit that learning is inherently social and knowledge does not belong to individual's mind. Therefore, learning is evidenced in group-level achievement, such as co-regulated actions, group cognition or knowledge advancement (Järvelä & Hadwin, 2013; Scardamalia & Bereiter, 2021; Stahl, 2016). Individual variability in participation is usually overlooked. Group-level success becomes the main unit of analysis, and sometimes, the endpoint of analysis, masking the diverse individual trajectories that underlie it. Third, the learning ecology perspective attempts to reconcile these views by treating individuals and collectives as interdependent, and learning occurs through their coordination and co-development (Damșa et al., 2019; White, 2018). However, many studies still examine individual learning after group work, rather than the transitions linking them. The challenge remains to explain how individual actions become part of collective work, and how collective achievements transform personal understanding.

Across these approaches, the persistent gap lies in the lack of systemic theorization of transitions across levels in CSCL. Existing models describe learning within levels or eventually reduce cross-level learning to one level. Without a theoretical framework to capture these transitions, design and analysis remain limited: educators may observe uneven participation or divergent learning outcomes but lack conceptual tools to interpret or support effectively. This motivates a framework that theorizes the relational, iterative, and mediated learning across levels.

Transitional Space: A conceptual framework

Theoretical grounding

Learning across levels is inherently relational (Lawrence & Valsiner, 1993; Stahl, 2016). Vygotskian sociocultural theory provides the foundation for this stance by locating learning as originating in social interaction and becoming internalized through participation in shared activity (Vygotsky, 1978). Higher mental functions first emerge between people on the inter-psychological plane before being reconstructed within the individual, on the intra-psychological plane. This view foregrounds interaction, language and cultural tools as mediators of learning: the mechanisms through which collective meaning-making becomes individual understandings (Bakhtin, 2010; Wertsch, 2001). Yet Vygotskian also implied reciprocity in this movement: individual development, once achieved, re-enters the social plane as learners contribute new ideas and practices, reshaping the collective environment from which their learning emerged. In summary, it reveals internalization and externalization as intertwined and continuously negotiated across levels of activity. Development occurs not within the boundaries of individuals or groups alone but through the transition that connects them to the spaces where social interaction and personal understanding intersect. This theoretical stance forms the foundation for the framework proposed here. By building on this perspective, *Transitional Space* captures the conditions, mechanisms, and supports that make such movement across levels observable and analyzable in a collaborative learning environment.

 © ISLS

Theorizing the framework: *Transitional Space*

We define it as a condition that an individual learner encounters and works with ideas generated in collaboration to refine their own thinking when students are in the discourse of the collaborative work. It can occur whenever a learner engages with group's ideas during collective discussion or when the learner independently synthesizes the group's contributions in a subsequence activity. It captures the relational process that links collective knowledge advancement with individual sense-making, emphasizing that learning occurs not solely within social interaction or internal cognition but in the movement between them (Rogoff et al., 1995; Wells, 2015). *Transitional Space* therefore provides a lens for understanding how learners traverse the boundaries of between collective and individual levels: how collective ideas are externalized, taken up and internalized in diverse ways in the discourse.

Enacted *Transitional Space* (interactional)

It responds to a central limitation in existing approaches to collaborative learning: the account of how collective progress becomes available for individual learning in the very moments when it occurs. It brings this invisible process into view by capturing the interactional junctures in which group reasoning advances beyond the understanding of some participants, creating openings for new forms of participation and sense-making. In these moments, the collective generates conceptual resources that individuals can draw on, reinterpret, or resist, and the resulting exchanges determine whether group level advancement contributes to individual transformation. In this way, enacted *Transitional Space* provides a theoretical and analytical means to observe the movement learning across levels as it unfolds in real time, rather than inferring it from outcomes after collaboration ends. It has two defining features that together explain how such cross-level movement becomes available.

First, it is enacted because it does not pre-scripted in the design or structure of the task but emerges dynamically from collaboration. When individuals encounter understandings that challenge their current understanding and invites reorganization of their thinking as a result of the group's discourse, a zone of possibility for transitions is created (Cang et al., 2024). Within this process, articulation and uptake play a central role. Articulation occurs as individuals externalize their thinking, through talk or contribution to shared artifacts, making internal processes and ideas accessible to others and expanding the group's joint knowledge constructing (Scardamalia & Bereiter, 2021). Uptake occurs as participants respond to, adapt or extend these articulated ideas, appropriating them in ways that connect collective insights with personal understanding (Stahl, 2013). The interplay of articulation and uptake thus makes the transitions observable through these cycles that ideas circulate, evolve and bridge the space between group-level meaning and individual sense-making. Second, it is interactional because the existence and accessibility of this space depend on how participants relate to one another. Some learners enter the discourse actively, integrating group ideas into their own reasonings, others remain peripheral or even disengaged, limiting their ability to benefit from collective advancement (Cang et al., 2024). The interactional patterns, which are revealed by the reciprocal moves of articulation and uptake, determine who can access the evolving collective understandings and how deeply they can engage with it (Lawrence & Valsiner, 1993). Both features indicates that *Transitional Space* is fragile and contingent. It arises only under specific discursive and social conditions that are often in unpredictable ways (Damşa et al., 2019). Its presence depends on the deviate or alignment between collective knowledge construction and individual participation trajectories (see another paper by Cang et al., 2026 in the same volume of this proceeding). Recognizing this fragility clarifies why collaborative learning cannot be assumed to produce reciprocal benefit: learning occurs when individuals successfully traverse enacted space, not simply by being present in groups.

Supported *Transitional Space* (design)

As the second facet, it complements the emergent nature of *Enacted Transitional Space* by emphasizing the intentional design conditions that make cross-levels learning possible and sustainable. While enacted dimension captures the spontaneous emergence of opportunities for learning within interactions discourse, *Supported Transitional Space* addresses a different but equally critical challenge, how to create environments enabling learners to reorganize, revisit and work productively with these chances. Collaborative learning does not automatically produce such transitions. Even when group work generates collective advances, learners may not have the time, structure, or tools to engage deeply with them (Celepkolu et al., 2017). *Supported Transitional Space* responds to this gap by focusing on how design features (, including tasks, scaffolds and representational resources, shape the conditions through which collective ideas can become personally meaningful.

It is featured by, first, supported because transitional processes rarely arise or persist on their own productively. Without structured opportunities to re-engage with group ideas, learners may leave collaborative activity without integrating the collective knowledge they helped generate. It thus depends on carefully designed scaffolds that bridge the gap between collective and individual understanding (Puntambekar, 2015; 2022). These scaffolds may take the form of sequenced activities that alternate between group and individual work, prompts

that encourage reflection on shared ideas, or tools that preserve and visualize group reasoning for later use (Zeng et al., 2024). Moreover, within these designed conditions, the processes of articulation and uptake that drive *Enacted Transitional Space* are intentionally extended and maintained. Opportunities for articulation, such as prompts to explain claims, annotate group products, or summarize evolving ideas, help learners externalize and refine their thinking after the group interaction. Opportunities for uptake, such as reflection tasks, comparison activities, or revisiting shared artifacts, encourage learners to interpret, adapt, and sense-making the group's collective work at their own pace. The temporary moments of understanding can be transformed into continued participation with shared knowledge, allowing individuals to advance and personalize what was collectively constructed (Cang et al., 2024). These designed features function also serve as mediating structures that make collective understanding traceable, accessible, and revisitable over time (Tissenbaum & Slotta, 2019), which is second key feature that create condition for ongoing articulation and uptake across levels of activity.

In this sense, *Supported Transitional Space* addresses the fragibility of enacted dimension by providing external conditions that support the movement of knowledge across levels. Whereas *Enacted Transitional Space* emerges unpredictably through discourse, *Supported Transitional Space* embeds opportunities for transition within the design of the learning environment. It stabilizes collective advances while creating new zones for articulation and uptake, enabling learners to return to and work with group ideas until they become personally meaningful. If *Transitional Space* is theorized only interactionally, then its existence is confined to observable high-quality group talk and participation. But this misses how unevenly individuals may benefit: some remain peripheral, unable to internalize. Conversely, if *Transitional Space* is framed only as a matter of design, the analysis risks overlooking the emergent dynamics of group discourse itself: one cannot scaffold a process without knowing when and how it arises spontaneously in interaction. Each facet reveals dimensions of the collaborative learning situation, providing a focused lens on aspects that are most relevant to the corresponding perspective. Together, these two facets foreground *Transitional Space* as both an emergent property of interaction and a product of design, providing a lens for analyzing and supporting how learners traverse the levels in learning.

Discussion and conclusion

Although few learning environments have been created explicitly to support *Transitional Space*, research provides insights into how particular design features enable across-level learning. Several guiding principles can be inferred to support transitions: (1) allow continuity and revisitation of previous ideas (Zhang et al., 2018), (2) enable visibility and accessibility of group knowledge (Schnabert & Bodemer, 2019), (3) enable reflection and interpretation of shared knowledge into personal understanding (Strauβ et al., 2024), and (4) integrate individual and collective activities (Celepkolu et al., 2017). These inferred principles do not prescribe fixed features but highlight how learning environments can be designed to make the transitions identified in *Transitional Space* more durable, observable, and actionable (see another paper .

In facilitating for cross-level learning, teachers may face additional challenges in guiding the temporal flow and interactional dynamics of collaborative learning. Some groups advance rapidly while others stall, and connections between levels are often left implicit (Olsen et al., 2021). Discourse quality varies widely: students may engage in procedural talk or remain hesitant to challenge peers (Clark, 2015). Decisions on when and how to intervene without interrupting its emergence are inherently challenging. From a *Transitional Space* standpoint, such orchestration problems arise because transitions themselves are under designed pedagogically. Instruction typically emphasizes maintaining talk within levels rather than linking them meaningfully across levels. When collaborative work moves forward, its potential to inform subsequent learning is often lost due to a lack of explicit bridging mechanisms. Studies point to sone strategies for design: flexible collaboration scripts can support coordination without over-regulating interaction (Strauß et al., 2024), awareness tools help teachers identify moments of imbalance or conceptual progress (Janssen & Bodemer, 2013). Yet these tools are most effective when coupled with *Transitional Space* thinking: when teachers treat the support of transitions as learning opportunities rather than logistical shifts. For example, a teacher might pause group discussion when new reasoning emerges, label the idea and prompt reflection before moving on (Clarke, 2015). These small interventions stabilize enacted *Transitional Space* and transform them into supported opportunities for continued learning. However, *Transitional Space* cannot by itself resolve all orchestration demands. Timing and sensitivity to students' affective and disciplinary contexts remain matters of teacher decision. It provides a conceptual lens for noticing and designing transitions but does not substitute for the improvisational expertise required to facilitate meaningful discourse.

Future work should empirically trace how transitional moments unfold, examine how different designs support their uptake, and explore how tools and instruction together support these cross-level movements. By treating learning as a continuous process that flows between social interaction and personal understanding, *Transitional Space* aims to provide a coherent foundation for linking theory, method and design in CSCL.

References

Bakhtin, M. M. (2010). *The dialogic imagination: Four essays.* University of Texas Press.

Barron, B. (2003). When smart groups fail. *The journal of the learning sciences, 12*(3), 307-359.

Cang, X., Zeng, L., Tissenbaum, M., Gnesdilow, D., & Puntambekar, S. (2024). Convergent Group Understanding: Indicator of Knowledge Integration?. In *Proceedings of the 17th International Conference on Computer-Supported Collaborative Learning-CSCL 2024, pp. 197-200*. International Society of the Learning Sciences.

Celepkolu, M., Wiggins, J. B., Boyer, K. E., & McMullen, K. (2017). Think first: Fostering substantive contributions in collaborative problem-solving dialogues. *Philadelphia, PA: International Society of the Learning Sciences.*

Chi, M. T. H. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, 1, 73-105. doi:10.1111/j.1756-8765.2008.01005.x

Clarke, S. N. (2015). The right to speak. *Socializing intelligence through academic talk and dialogue,* (s 167), 180.

Damşa, C., Nerland, M., & Andreadakis, Z. E. (2019). An ecological perspective on learner-constructed learning spaces. *British Journal of Educational Technology, 50*(5), 2075-2089.

Hakkarainen, K. (2009). A knowledge-practice perspective on technology-mediated learning. *International Journal of Computer-Supported Collaborative Learning, 4*, 213-231.

Järvelä, S., & Hadwin, A. F. (2013). New frontiers: Regulating learning in CSCL. *Educational psychologist, 48*(1), 25-39.

Lawrence, J. A., & Valsiner, J. (1993). Conceptual roots of internalization: From transmission to transformation. *Human development, 36*(3), 150-167.

Olsen, J. K., Rummel, N., & Aleven, V. (2021). Designing for the co-orchestration of social transitions between individual, small-group and whole-class learning in the classroom. *International Journal of Artificial Intelligence in Education, 31*(1), 24-56.

Pea, R. D. (1993). Practices of distributed intelligence and designs for education. *Distributed cognitions: Psychological and educational considerations, 11*, 47-87.

Puntambekar, S. (2015). Distributing scaffolding across multiple levels: Individuals, small groups, and a class of students. Essential readings in problem-based learning, 207-221.

Puntambekar, S. (2022). Distributed scaffolding: Scaffolding students in classroom environments. *Educational Psychology Review*, 34(1), 451-472.

Rogoff, B., Baker-Sennett, J., Lacasa, P., & Goldsmith, D. (1995). Development through participation in sociocultural activity. *New directions for child and adolescent development, 1995*(67), 45-65.

Scardamalia, M., & Bereiter, C. (2021). Knowledge building: Advancing the state of community knowledge. *International handbook of computer-supported collaborative learning*, 261-279.

Schnaubert, L., & Bodemer, D. (2022). Group awareness and regulation in computer-supported collaborative learning. *International Journal of Computer-Supported Collaborative Learning, 17*(1), 11-38

Stahl, G. (2016). From intersubjectivity to group cognition. *Computer Supported Cooperative Work (CSCW), 25*, 355-384.

Stahl, G. (2013). A model of collaborative knowledge-building. *In International conference of the learning sciences* (pp. 70-77). Psychology Press.

Strauß, S., Tunnigkeit, I., Eberle, J., Avdullahu, A., & Rummel, N. (2024). Comparing the effects of a collaboration script and collaborative reflection on promoting knowledge about good collaboration and effective interaction. *International Journal of Computer-Supported Collaborative Learning*, 1-39.

Tissenbaum, M., & Slotta, J. D. (2019). Developing a smart classroom infrastructure to support real-time student collaboration and inquiry: a 4-year design study. *Instructional Science, 47*, 423-462.

Vygotsky, (1978). *Mind in Society: Development of Higher Psychological Processes* (M. Cole, V. Jolm-Steiner, S. Scribner, & E. Souberman, Eds.). Harvard University Press. https://doi.org/10.2307/j.ctvjf9vz4

Wells, G. (2015). Dialogic learning: Talking our way into understanding. *Education as Social Construction*, 62.

Wertsch, J. (2001). The multivoicedness of meaning. Discourse theory and practice, 222-235.

White, T. (2018). Connecting levels of activity with classroom network technology. *International journal of computer-supported collaborative learning, 13*, 93-122.

Zeng, L., Cang, X., Sultana, A., Cernova, S., Wang, M., Gnesdilow, D., ... & Tissenbaum, M. (2024). Spatial manipulative note-taking tool for small group face-to-Face collaboration in science class. *Proceedings of the 17th International Conference on Computer-Supported Collaborative Learning-CSCL 2024*.

Zhang, J., Tao, D., Chen, M. H., Sun, Y., Judson, D., & Naqvi, S. (2018). Co-organizing the collective journey of inquiry with idea thread mapper. *Journal of the Learning Sciences, 27*(3), 390-430.

International Society of the Learning Sciences

From Words to Networks: AI-Powered Knowledge Graphs for Understanding Collaborative Learning

Ahmad Khanlari, University of Toronto, a.khanlari@mail.utoronto.ca

Abstract: This paper presents an analytics tool that leverages knowledge graphs to make visible the complex, collaborative discourse in Knowledge Forum. The tool extracts relational triplets (subject, predicate, object) from student posts using a hybrid pipeline combining LLM-based, transformer-based (Triplex), and rule-based (spaCy) methods. Extracted triplets are normalized and visualized as interactive graphs, where nodes represent concepts and edges represent relationships. A case study on the salmon life cycle demonstrates how the visualization surfaces central ideas, thematic clusters, and knowledge gaps, aligning with key Knowledge Building principles. By transforming unstructured discourse into an explorable network, the tool supports students and educators in tracing idea development and identifying opportunities for deeper inquiry.

Introduction

Knowledge Building, as proposed by Scardamalia and Bereiter (2006), is a pedagogical approach that prioritizes collaborative discourse to create knowledge and advance collective understanding. Knowledge Forum, a digital platform designed to support collaborative knowledge creation, enables students to post, discuss, and refine ideas through text-based contributions (Scardamalia, 2004). However, analyzing the complex, interconnected nature of student discourse to assess knowledge advancement remains a challenge for educators. This paper introduces a novel knowledge graph-based analytics tool that transforms student discourse into structured, visual representations, making the progression of ideas explicit and actionable.

The concept of knowledge graphs was popularized by Google's Knowledge Graph in 2012, which shifted search from keyword matching to entity-based understanding (Uyar & Aliu, 2015). Knowledge graphs are distinct from other data representations because they integrate information into an ontology and apply reasoning to derive new knowledge (Uyar & Aliu, 2015). These knowledge graphs represent information as a network of entities (nodes) connected by relationships (edges), typically in the form of triplets (subject, predicate, object), enabling complex queries and reasoning (Ehrlinger & Wöß, 2016). For example, in the sentence "Salmon live in the Humber River," the triplet (Salmon, lives in, Humber River) captures the relationship (i.e., lives in) between the entities "Salmon" and "Humber River." These graphs emphasize semantic relationships, allowing for flexible integration of diverse data sources and supporting advanced analytics. Knowledge graphs have gained prominence in various fields, including artificial intelligence and data integration, due to their ability to model complex, interconnected information (Ehrlinger & Wöß, 2016).

In education, knowledge graphs are increasingly used to represent and analyze structured knowledge, such as in personalized learning, curriculum design, and concept mapping (Abu-Salih & Alotaibi, 2024). Prior work has constructed knowledge graphs through expert-curated resources or automated pipelines relying on single-method approaches, including rule-based or machine learning techniques. While these methods are effective in controlled settings, they often struggle with informal, diverse student discourse, leading to trade-offs between precision, coverage, and adaptability.

Our approach builds on these traditions by integrating complementary paradigms within a unified pipeline. By combining LLM-based, transformer-based, and rule-based methods, the system leverages their respective strengths to produce a more robust and semantically rich representation of discourse than single-method approaches. Beyond methodological integration, this tool extends existing Knowledge Building analytics by transforming discourse into an explicit network of relationships, rather than relying solely on frequency-based or note-level analyses. This enables the identification of idea connections, gaps, and misconceptions, making the structure and evolution of collective knowledge more visible and actionable.

Triplet extraction

Knowledge graphs can serve as a tool to visualize the collective knowledge constructed by the community, revealing how students' ideas are connected and evolve over time. To construct knowledge graphs from student discourse, our tool extracts relational triplets (subject, predicate, object) using three complementary methods. These methods are applied to capture different types of linguistic patterns in student writing, improving the completeness and reliability of extracted relationships. Below, we describe each method and its role in the pipeline.

 © ISLS

Method 1: LLM-based triplet extraction

We leverage a transformer-based Large Language Model (LLM) to automatically extract relational triplets—subject, predicate, object—from unstructured student discourse. LLMs, built on parallelized attention mechanisms, excel at capturing long range dependencies and nuanced contextual cues in text (Vaswani et al., 2017). This makes LLMs particularly suited to identify diverse linguistic constructions and implicit semantic links that simpler, rule-based extractors often miss. For our implementation, we employ *Zephyr-7B*, an open-source GPT-style language model developed by Hugging Face.

Such LLM-based extraction offers substantial flexibility: it readily adapts to varied sentence structures, captures subtle semantic relations (e.g., cause–and–effect, part–whole), and is robust against noisy or colloquial language typical of student discourse (Ateia et al., 2025). However, its performance depends critically on prompt clarity and can incur higher compute costs compared to rule-based methods. To mitigate these challenges, we combine LLM outputs with complementary extraction strategies (see Methods 2 and 3), ensuring both precision and coverage in our knowledge graph construction.

Method 2: Transformer-based triplet extraction with Triplex

To complement our LLM-based approach, we employ Triplex, a transformer-based model explicitly optimized for knowledge graph construction. Whereas general-purpose LLMs excel at language understanding across diverse tasks, transformers like Triplex are fine tuned to detect structured relations and domain-specific entities, delivering higher precision and efficiency for large-scale discourse analysis (Chen et al., 2023). Triplex, like other transformers, leverages self-attention mechanisms to process all tokens in a text sequence in parallel, modelling long-range dependencies and complex contextual interactions (Vaswani et al., 2017). These properties make transformer architectures a reliable tool for tasks like Named Entity Recognition (NER) and relationship extraction, where precise identification of entity boundaries and their links is critical (Eberts & Ulges, 2019).

Through fine-tuning, Triplex learns to:

1. Recognize predefined entity types (e.g., animal, location, life stage) with high recall and precision.
2. Classify relations using a fixed ontology of predicates (e.g., lives in, eats, is part of), ensuring extracted triplets align with our domain taxonomy.
3. Handle noisy text, such as informal student posts, through robust tokenization and context-aware inference.

Triplex's domain specialization affords faster inference and higher extraction accuracy compared to more general LLMs. By limiting its relation vocabulary, it reduces noise and ensures consistency with our knowledge graph schema (Chen et al., 2023). When combined with the flexibility of LLM-based extraction and the precision of rule-based patterns (Method 3), Triplex helps us achieve a balanced pipeline that maximizes coverage while controlling false positives.

Method 3: POS tagging and dependency parsing with spaCy

For this method, we used spaCy, which is a high performance, open source Python library offering non-destructive tokenization, part-of-speech (POS) tagging, and dependency parsing, among other NLP components (Honnibal et al., 2020). POS tagging assigns each token a grammatical category (noun, verb, adjective, etc.), providing the foundation for identifying core sentence elements like subjects and objects (Meftah et al., 2018). We begin by preprocessing each student's note with spaCy's sentence boundary detection, segmenting the discussion into coherent sentences before normalizing punctuation and expanding contractions to enhance tagging accuracy. To focus on factual statements, the tool then filters questions by analyzing linguistic features, including:

- Wh-words: Words like "what," "where," or "how" that often signal questions.
- Auxiliary Verbs: Sentences beginning with verbs like "Do," "Is," or "Are."
- Sentence Structure: Patterns where the main verb is linked with a subject and auxiliary verb, typical in questions (e.g., "Is salmon black?").

Each sentence is then run through a lightweight English language model provided by spaCy, to produce token-level part-of-speech tags and a full dependency parse tree that reveals the grammatical structure of the sentence—such as which words are subjects, objects, modifiers, and nominal subject (nsubj) of a verb that, in turn, governs either a direct object (dobj) or a prepositional phrase (prep) introducing another noun. For instance, spaCy parses "Salmon swim up rivers" by marking "salmon" as the nsubj of "swim" and linking "rivers" through a prep→pobj chain, yielding the triplet (salmon, swim up, rivers). SpaCy's rule-based method provides high precision on grammatical constructions, effectively capturing triplets from standard sentences.

 © ISLS

Integration and visualization

The pipeline preprocesses and segments student notes, followed by parallel triplet extraction using LLM-based, Triplex, and spaCy methods. Extracted triplets are then normalized, deduplicated, and merged into a unified dataset through rule-based aggregation and conflict resolution. Each method contributes complementary strengths—spaCy provides syntactic precision, Triplex ensures schema consistency, and LLM-based extraction captures implicit and context-dependent relations—resulting in a balanced approach that improves coverage, accuracy, and robustness. In this integrated framework, spaCy anchors the graph in structurally precise relations, while the other methods enrich it with semantic and domain-specific connections, enabling a more comprehensive representation of idea progression in Knowledge Forum.

The resulting dataset is stored in a central CSV file and visualized using Pyvis as an interactive graph, where nodes represent concepts and edges represent relationships. Edge types are color-coded, and users can explore the graph by adjusting layouts, inspecting source sentences, and filtering relations, supporting the identification of idea pathways, central concepts, and under-explored areas in student discourse.

Visual example

To illustrate the tool's application, a knowledge graph created from student discourse in Knowledge Forum on the topic of salmon is shown in Figure 1. This knowledge graph visualizes the semantic relationships derived from student discourse on the salmon lifecycle. The graph consists of nodes representing key concepts (e.g., "Salmon," "Egg," "Fry," "Ocean," "River") and edges indicating relationships (e.g., "lay eggs in," "grow in," "eaten by"). The graph highlights central concepts and reveals a hierarchical and interconnected structure of knowledge. The "Salmon" node serves as a hub, with radiating connections to lifecycle stages, environmental factors ("River," "Ocean," "Waterfall"), and biological interactions ("Eaten by Bear," "Eaten by Eagle").

The central node, "Salmon," also connects to multiple stages of the lifecycle ("Egg," "Fry," "Smolt," "Adult"), reflecting the cyclical nature of the topic and the students' understanding of its progression through stages (labeled as Stage 1 to Stage 4). This centrality underscores "salmon" as the focal point of discourse, while the density of connections around lifecycle stages (e.g., "Egg" to "Fry" to "Smolt") indicates a strong grasp of sequential development. Notably, the graph also captures external influences, such as predators ("Bear," "Eagle") and human-related factors ("Fish Hook," "People"), suggesting that students are integrating ecological and anthropogenic dimensions into their understanding.

The graph also highlights patterns in knowledge construction. For instance, the cluster around "Egg" (connected to "Lay eggs in," "Hatch," "River") shows that students associate early lifecycle stages with specific environmental contexts. Similarly, the "Ocean" node connects to "Smolt" and "Adult," reflecting an understanding of habitat transitions. However, some nodes, such as "Adipose Fin" and "Salmon Jump," are less connected, indicating potential gaps in discourse or areas where students' knowledge is less developed. These peripheral concepts may require targeted instructional strategies to better integrate them into the broader conceptual framework.

Implications and discussion

The knowledge graph provides actionable insights into student discourse by making relationships between ideas visible. Prominent connections (e.g., *Salmon → River → Egg*) indicate strong foundational understanding, while sparsely connected concepts (e.g., *Adipose Fin*) highlight opportunities for targeted scaffolding. By surfacing misconceptions (e.g., incorrect relations), the graph supports peer critique and iterative refinement, enabling a more transparent and collective knowledge building process. It also facilitates structured classroom activities, such as identifying gaps, guiding discussion, and supporting data-informed instructional decisions.

By representing discourse as a network, the tool allows educators and students to identify central ideas, emerging themes, and underdeveloped concepts, providing a holistic view of learning. The integration of LLM-based, transformer-based, and rule-based extraction methods enables both broad semantic coverage and structural precision, supporting robust analysis of collaborative inquiry.

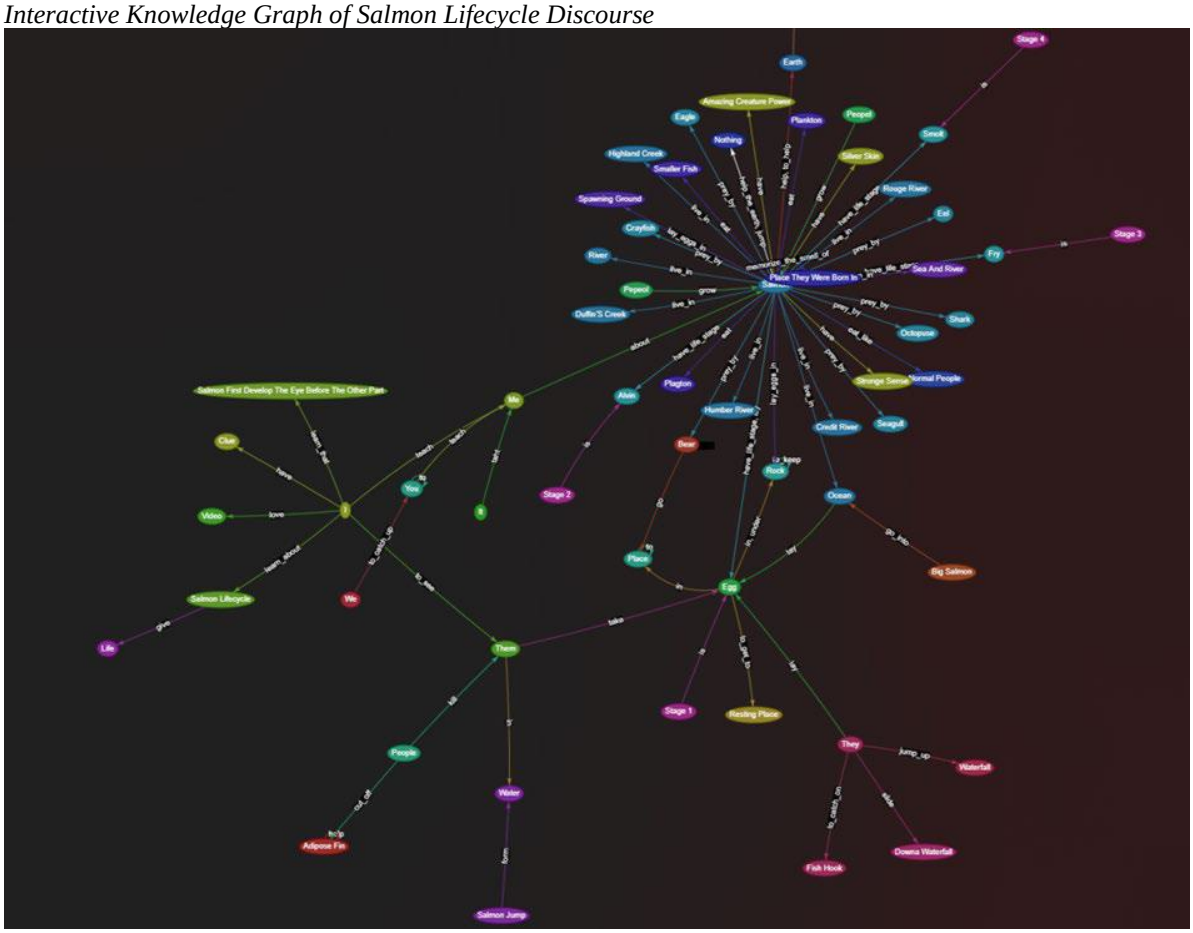
Grounded in Knowledge Building theory, the graph operationalizes key principles in a concrete and visible way. It supports *Idea Diversity* by representing each distinct contribution as a node, allowing multiple perspectives and ideas to coexist and be explored. *Community Knowledge* emerges as individual contributions are aggregated into shared structures, making collective understanding explicit. The principle of *Rise Above* is facilitated by revealing higher-level conceptual patterns and connections across ideas, encouraging learners to synthesize and move beyond isolated facts. *Symmetric Knowledge Advancement* is supported by enabling students to view, build on, and refine one another's contributions through the evolving network of relationships. In this

 © ISLS

way, the tool functions as both an analytic lens and a pedagogical scaffold, making the dynamics of collective knowledge construction visible and actionable for learners and educators.

Figure 1

Interactive Knowledge Graph of Salmon Lifecycle Discourse



The empirical example presented here is illustrative, demonstrating the analytic affordances of the approach rather than providing a full evaluation. Accordingly, this work is best understood as a design-oriented contribution. Despite its potential, classroom application depends on effective teacher facilitation and integration into lesson plans, underscoring the importance of involving teachers as co-designers. Teacher perspectives and classroom-based feedback were not systematically included, representing a key limitation. Future work will incorporate co-design with teachers, qualitative user studies across diverse contexts, and evaluation of both technical validity (e.g., reliability and coverage of extracted triplets) and pedagogical impact. Further development will also focus on improving adaptability across topics and supporting flexible, context-sensitive use in applied settings.

Conclusion

This study demonstrates the feasibility and educational promise of leveraging knowledge graphs to analyze and support Knowledge Building discourse. By automatically extracting relational triplets from student posts and rendering them as interactive networks, the tool makes the invisible architecture of collaborative inquiry explicit and actionable. Students and educators gain real time insights into how ideas emerge, intersect, and evolve—insights that were previously obscured by the volume and complexity of textual contributions. The integration of multiple NLP methods ensures robustness across varied linguistic styles, while the Pyvis visualization offers an intuitive interface for both teacher and learner engagement. Ultimately, by coupling advanced graph analytic techniques with foundational Knowledge Building theory, this research paves the way for smarter, more responsive learning environments that transform communal idea advancement from an implicit promise into an explicit, navigable reality.

 © ISLS

References

Abu-Salih, B., & Alotaibi, S. (2024). A systematic literature review of knowledge graph construction and application in education. *Heliyon*, *10*(3).

Ateia, S., Kruschwitz, U., Scholz, M., Koschmider, A., & Almohaishi, M. (2025). LLM-based information extraction to support scientific literature research and publication workflows. In *New Trends in Theory and Practice of Digital Libraries: TPDL 2025* (pp. 90–99). Springer. https://link.springer.com/chapter/10.1007/978-3-032-06136-2_9

Chen, S., Cheng, H., Liu, X., Jiao, J., Ji, Y., & Gao, J. (2023). Pre-training transformers for knowledge graph completion. *arXiv preprint arXiv:2303.15682*. https://doi.org/10.48550/arXiv.2303.15682

Eberts, M., & Ulges, A. (2019). *Span-based Joint Entity and Relation Extraction with Transformer Pre-training (SpERT).* In *Proceedings of the 24th European Conference on Artificial Intelligence* (pp. 2103–2110).

Ehrlinger, L., & Wöß, W. (2016). Towards a definition of knowledge graphs. In *Joint Proceedings of the Posters and Demos Track of the 12th International Conference on Semantic Systems (SEMANTiCS 2016) and the 1st International Workshop on Semantic Change & Evolving Semantics (SuCCESS16)* (Vol. 1695). CEUR-WS.

Scardamalia, M., & Bereiter, C. (2006). *Knowledge building: Theory, pedagogy, and technology*. In K. Sawyer (Ed.), Cambridge Handbook of the Learning Sciences (pp. 97-118). New York: Cambridge University Press.

Uyar, A., & Aliyu, F. M. (2015). Evaluating search features of Google Knowledge Graph and Bing Satori: Entity types, list searches and query interfaces. *Online Information Review*, *39*(2), 197-213

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*.

 © ISLS

International Society of the Learning Sciences

How Active User Count Relates to Group Outcomes in Collaborative Programming: An Implicit Scripts Perspective

Zhenyu Cai, EPFL, zhenyu.cai@epfl.ch
Richard Lee Davis, KTH Royal Institute of Technology, rldavis@kth.se
Roland Tormey, EPFL, roland.tormey@epfl.ch
Pierre Dillenbourg, EPFL, pierre.dillenbourg@epfl.ch

Abstract: Collaboration scripts structure how learners work together by prescribing roles, tasks, and expected contributions. When scripts are implicit rather than explicit, learners decide how to enact them, which can shift the number of active participants and potentially outcomes. Prior work suggests sequential tasks may not benefit from multiple simultaneous actors due to the misalignment with the implicit script. Building on this, we test how the number of active users relates to time-on-task and group performance. Using panel data from three sequential Jupyter-notebook assignments, we estimate within-group fixed effects models to examine early-stage and aggregated activity signals. We find that active user count is significantly associated with both outcomes, while early signals show a modest correlation with group scores.

Introduction and background

While group size remains a visible design variable in collaborative learning, evidence converges on the view that patterns of interactional dynamics are the key determinants. Larger groups often introduce coordination costs that depresses engagement and motivation, inviting social loafing (Latané et al., 1979) or free riding (Kerr & Bruun, 1983). However, headcount alone is a weak predictor of outcomes, if not taking into consideration the contextual factors like task structures or technical constraints. What truly matters is how many members are actively engaged within a shared problem space and how their contributions converge meaningfully (Dillenbourg, 1999)—through transactive talk, turn-taking, mutual regulation, etc. (Barron, 2003). In short, the effective size of a group is the subset of members, the active learners, who sustain task-focused, responsive participation.

In CSCL, small groups can look very different. Some dyads or triads co-construct on a shared screen or document with continuous discourse; others "split and merge", parallelizing work with minimal communication. A key differentiator is the alignment between pedagogical task design with the tool's affordances. If the task doesn't require interdependence—or the technology resists it—collaboration collapses. CSCL tools that provide group awareness, or what Suthers (2006) calls "representational guidance", such as shared visual references, are therefore essential. Well-designed environments often make collaboration feel natural and effortless.

Beyond tools, collaboration scripts address recurring coordination and participation problems by specifying roles, phases, and expected contributions. Scripts convert mere headcount into functional interdependence, increasing the likelihood that learners engage with each other rather than silently parallelize (Kollar et al., 2006). Prior distinctions between micro-scripts and macro-scripts clarify how scripts can reduce coordination costs while fostering transactivity, grounding, and regulation—mechanisms that matter more than size per se for learning (Dillenbourg, 2002; Fischer et al., 2013). Some scripts are intrinsically dyadic because they rely on complementary roles. The ArgueGraph script, for instance, pairs learners with opposing positions, and guides them through a cycle of comparing, discussing, and reconciling views (Jermann & Dillenbourg, 2003).

Recent work introduces *implicit scripts*—structural affordances embedded in learning materials that guide collaboration without explicit instruction (Cai, Davis, Tormey, & Dillenbourg, 2025). In computational notebooks, graph-based methods can operationalize task dependencies and surface teamwork strategies. Notably, for notebooks that lack real-time co-editing capabilities but possess sequential dependencies, the effective strategy observed was to open a single notebook and have the team collaborate on one screen—yielding one "active user" at a time. While active user counts helped verify groups' adherence to the implicit script, the study did not examine whether this approach improved outcomes, leaving a critical question about the pedagogical value of such alignment. In response to this, we treat this dimension as a latent factor of implicit script adherence and explore its association with time-on-task and group performance.

Time-on-task

Time-on-task reflects how much time learners actually engage with learning activities. It has long been used in educational research to estimate learning effort, assess instructional design, and explain outcome variance (Stallings, 1980). In learning analytics, time-on-task is typically paired with frequency measures (e.g., event counts), although it might be operationalized differently depending on the types of trace data. In this regard,

 © ISLS

Kovanovic et al. (2015) compared time-on-task estimation strategies and suggested careful method selection. Following recent practice (Gao et al. 2025), we use wall-clock time and active time as our estimates.

Research questions

In this paper, we address a gap in work on implicit scripts by testing how the number of active users relates to group outcomes in collaborative programming. We replicate prior settings with panel data from three non-collaborative, sequential Jupyter notebooks—the same context in which an implicit script encourages a single-notebook workflow. We estimate within-group associations using two-way (group × assignment) fixed effects (FE) regressions with group-clustered standard errors to partial out time-invariant factors. Accordingly:

- RQ1: To what extent does the number of active users in computational notebooks associate with group-level time-on-task and assignment performance?
- RQ2: Can early counts of active users serve as a practical indicator of subsequent time-on-task and group performance?

Methods

Participants and data collection

The study took place in a second-year undergraduate mathematics course for life-science majors at a European university. In addition to weekly in-person lectures, students had optional lab sessions using Jupyter notebooks. Five group assignments were submitted during these sessions across the semester. Within the university-hosted platform, there was no collaborative version of Jupyter notebooks, so student groups either worked on a single notebook on one device or used separate copies on different devices and merged them later.

Of the 187 enrolled students (46 groups: 43 with four members, the rest with five), 147 (78.61%) consented to data collection. Ethics approval was obtained from the university. Participation was voluntary and had no impact on grading. Groups were included in the analysis only if all members who generated log data had provided consent, resulting in 33 groups being analyzed.

An event tracking system (Cai, Davis, Mariétan, et al., 2025) was installed for collecting students' interaction data in Jupyter notebooks, including clicks on notebooks and cells, cell creation and removal, and cell execution with output and status. Group-level scores (0.00–1.05) for each assignment were also collected. To support reproducibility and replicability, the panel dataset, documentation, preprocessing scripts, and individual data (presence records and final exam grades) are available at https://doi.org/10.5281/zenodo.19316212.

Data analysis

Three sequentially structured assignments were retained for analysis: Assignment 2 (Lagrange Interpolation), Assignment 3 (High Order Quadrature Formula), Assignment 4 (Iterative Methods for Linear Systems). One group from Assignment 2 and two groups from Assignment 4 were excluded due to incomplete participation data (e.g., some team members used other programming platforms). Table 1 summarizes the group member presence records in each assignment session. All analyses were conducted using Python and R.

Table 1

Group Member Presence Records Across Three Assignments

	Frequency		
Presences	Assignment 2	Assignment 3	Assignment 4
N = 3	0	1	3
N = 4	31	30	28
N = 5	1	2	0
Total	32	33	31

N refers to physical attendance, which differs from the logged user count.

To address RQ1, we ran FE regressions using the number of active users—*user_count*, defined as the total number of users who executed at least one cell (rather than just clicked) within the time window—as the predictor, and the group-level time-on-task and assignment raw score as outcomes. Specifically, we employed two time-on-task metrics: *time_login* measured the interval between the first login by any group member and the assignment submission, representing the total duration of engagement; *time_exec* was derived from the first to the last timestamps of cell executions prior to submission, indicating the period of active programming work. Both group and assignment FEs were controlled to isolate the models from any systematic differences across groups or

 © ISLS

across assignments. Standard errors (SEs) were clustered at the group level to allow for arbitrary correlation of residuals within each group across assignments.

For RQ2, the dependent variables were determined based on the results from RQ1. We divided the 1.5-hour programming session into three 30-minute segments (08:30–09:00, 09:00–09:30, and 09:30–10:00) and counted the number of active users within each time window. Within-subject *t*-tests were conducted to compare these segmented measures with the previously aggregated one, and a bar plot was generated to visualize the overall changes. Subsequently, we used FE regression models to examine the associations between these factors and outcome variables, assessing their predictive potential by comparing model significance with the RQ1 results.

Results

RQ1: Association between active user count and time-on-task/performance

Table 2
Descriptive Statistics of Three Assignments (A2: N=32; A3: N=33; A4: N=31)

	user_count			*time_exec (s)*			*time_login (s)*			*raw_score*		
	M	SD	Md	M	SD	Md	M	SD	Md	M	SD	Md
A2	2.44	1.08	2.50	4250	641	4302	4699	553	4834	0.82	0.11	0.85
A3	2.58	1.15	2.00	4124	795	4288	4566	439	4498	0.91	0.11	0.95
A4	2.39	1.15	2.00	4349	783	4405	4762	519	4801	0.89	0.10	0.90

Table 2 summarizes the descriptive statistics of all variables across three assignments. Overall, these metrics were consistent across assignments, with only minor fluctuations in *time_exec* and *raw_score*. FE regressions were then employed to examine the association between *user_count* and the above three outcome variables. As shown in Table 3, with group and assignment FEs (31 groups, 3 assignments) and SEs clustered by group (N = 96), a higher *user_count* is associated with longer time-on-task and lower raw scores. Specifically, each additional user corresponds to an increase of 377.9 seconds in *time_exec* (SE = 168.1, $p < .05$; Adj. $R^2 = .441$; Within $R^2 = .105$; RMSE = 438.0), a non-significant increase of 103.0 seconds in *time_login* (SE = 72.9, $p \geq .05$; Adj. $R^2 = .501$; Within $R^2 = .020$; RMSE = 283.1), and a 0.066 decrease in raw_score (SE = 0.023, $p < .01$; Adj. $R^2 = .313$; Within $R^2 = .111$; RMSE = 0.075). Notably, the coefficients on *user_count* reflect how changes in the number of active users within a group across assignments—controlling for both group-specific and assignment-wide factors—relate to changes in both assignment performance and time-on-task (i.e., *time_exec*).

Table 3
FE Regressions on Three Outcome Variables

	time_exec (s)	*time_login (s)*	*raw_score*
user_count	377.9* (168.1)	103.0 (72.9)	-0.066** (0.023)
FEs	group (33), assignment (3)		
Observations	96		
Adj. R^2	0.441	0.501	0.313
Within R^2	0.105	0.020	0.111
RMSE	438.0	283.1	0.075

Standard errors clustered at the group level. *** $p < .001$, ** $p < .01$, * $p < .05$

RQ2: Active user count as an early indicator of time-on-task/performance

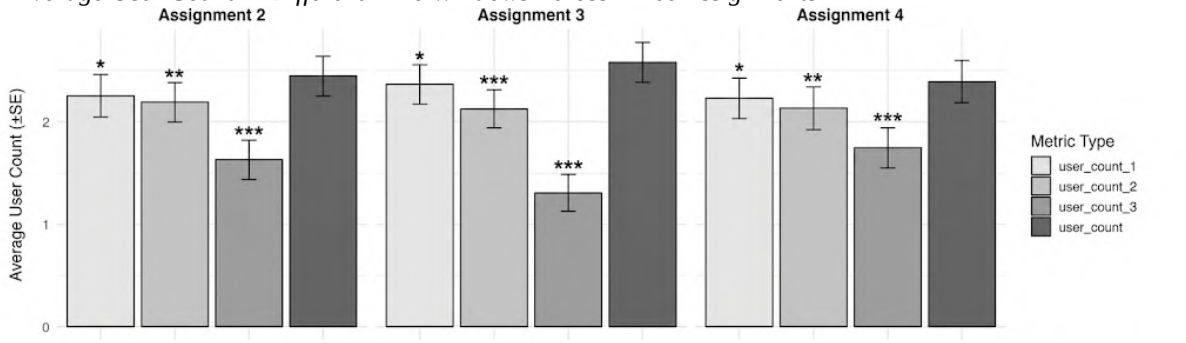
Based on the results from RQ1, *time_exec*, indicating active time-on-task, and *raw_score* remained as measures of group outcomes. To test whether *user_count* can function as an early indicator of these two outcomes, the 1.5-hour session was divided into three 30-minute intervals. The number of active users was recounted within each time window: *user_count_1*, *user_count_2*, and *user_count_3* correspond to the segmented *user_count* during 08:30–09:00, 09:00–09:30, and 09:30–10:00.

As shown in Figure 1, a consistent decline over time could be observed in the average *user_count* of all groups across all assignment, while the associated SEs remained relatively stable. The lowest value was the *user_count_3* in Assignment 3, which was nearly half of the aggregated one for that assignment. Within-subject *t*-tests confirmed that all segmented *user_count* were significantly different from the aggregated ones in each assignment, indicating clear temporal dynamics in group participation throughout the session.

 © ISLS

Figure 1

Average User Count in Different Time Windows Across Three Assignments



Within-subject *t*-tests were conducted between each segmented *user_count* and the aggregated *user_count* for each assignment. *** $p < .001$, ** $p < .01$, * $p < .05$

Table 4

FE Regressions of Segmented user_count on time_exec and raw_score

	time_exec (s)			*raw_score*		
user_count_1	130.2 (222.8)			-0.041* (0.018)		
user_count_2		351.7* (164.0)			-0.016 (0.022)	
user_count_3			517.6*** (102.4)			-0.020 (0.017)
FEs	group (33), assignment (3)			group (33), assignment (3)		
Observations	96			96		
Adj. R^2	0.382	0.441	0.629	0.255	0.234	0.244
Within R^2	0.010	0.105	0.406	0.035	0.008	0.022
RMSE	460.6	438.0	357.0	0.078	0.079	0.078

Standard errors clustered at the group level. *** $p < .001$, ** $p < .01$, * $p < .05$

Next, the associations between these segmented *user_count* and the group outcomes were examined using two-way FE models (group and assignment FEs; SEs clustered by group; N = 96). The results were summarized in Table 4. Specifically, *user_count_1* is negatively associated with *raw_score* ($\beta = –0.041$, SE = 0.018, $p < .05$)—about a 0.04 point drop in score per additional active user within a group—whereas *user_count_2* and *user_count_3* are small and not significant ($\beta = –0.016, 0.022$; $\beta = –0.020, 0.017$). In contrast, *time_exec* rises sharply with later time windows: with an increase of 351.7 seconds (SE = 164.0, $p < .05$) for *user_count_2* and 517.6 seconds (SE = 102.4, $p < .001$), with *user_count_3* explaining much of the within-assignment variation (Within $R^2 = 0.406$). Taken together, *user_count* in the early session (08:30–09:00) appeared to be a leading indicator of lower group performance, while in later sessions it reflected teamwork inefficiency (i.e., longer time-on-task) more than additional score losses.

Discussion

Building on earlier work on how learners collaborate on sequential tasks through the lens of implicit scripts, our study advanced the analysis by testing how the number of active users relates to group outcomes: time-on-task and assignment performance. Addressing RQ1 and RQ2, we found that: (1) lower concurrent active users are associated with higher group scores and less time-on-task; (2) this same metric, when measured in the first 30 minutes, provides an actionable predictive signal for these outcomes. This aligns with the prior observations (Cai, Davis, Tormey, & Dillenbourg, 2025): when the task is indivisible, the optimal strategy to prevent delays and underperformance is for the team to collaborate using a single, shared reference. Future work could operationalize early active user counts as a group performance indicator. In collaborative learning analytics (Wise et al., 2021), awareness tools and adaptable CSCL systems could surface this metric in real time—e.g., as a proxy for script adherence—to monitor group dynamics and trigger timely interventions.

 © ISLS

References

Barron, B. (2003). When smart groups fail. *The Journal of the Learning Sciences*, *12*(3), 307–359. https://doi.org/10.1207/S15327809JLS1203_1

Cai, Z., Davis, R. L., Mariétan, R., Tormey, R., & Dillenbourg, P. (2025). Jupyter Analytics: A toolkit for collecting, analyzing, and visualizing distributed student activity in Jupyter notebooks. In *Proceedings of the 56th ACM Technical Symposium on Computer Science Education V. 1* (pp. 172–178). ACM. https://doi.org/10.1145/3641554.3701971

Cai, Z., Davis, R. L., Tormey, R., & Dillenbourg, P. (2025). "Jupyter-Notebook-as-Script": Investigating the nature and impact of implicit collaboration scripts in computational notebooks. In J. Oshima, B. Chen, F. Vogel, & J. Järvelä (Eds.), *Proceedings of the 18th International Conference on Computer-Supported Collaborative Learning – CSCL 2025* (pp. 223–231). International Society of the Learning Sciences. https://repository.isls.org//handle/1/11847

Dillenbourg, P. (1999). What do you mean by collaborative learning? In P. Dillenbourg (Ed.), *Collaborative learning: Cognitive and computational approaches* (pp. 1–19). Elsevier.

Dillenbourg, P. (2002). Over-scripting CSCL: The risks of blending collaborative learning with instructional design. In P. A. Kirschner (Ed.), *Three worlds of CSCL: Can we support CSCL?* (pp. 61–91). Open Universiteit Nederland.

Fischer, F., Kollar, I., Stegmann, K., & Wecker, C. (2013). Toward a script theory of guidance in computer-supported collaborative learning. *Educational Psychologist*, *48*(1), 56–66. https://doi.org/10.1080/00461520.2012.748005

Gao, G., Leon, A., Jetten, A., Turner, J., Almoubayyed, H., Fancsali, S., & Brunskill, E. (2025). Predicting long-term student outcomes from short-term EdTech log data. In *Proceedings of the 15th International Learning Analytics and Knowledge Conference* (pp. 631–641). ACM. https://doi.org/10.1145/3706468.3706552

Jermann, P., & Dillenbourg, P. (2003). Elaborating new arguments through a CSCL script. In J. Andriessen, M. Baker, & D. Suthers (Eds.), *Arguing to learn: Confronting cognitions in computer-supported collaborative learning environments* (pp. 205–226). Springer. https://doi.org/10.1007/978-94-017-0781-7_8

Latané, B., Williams, K., & Harkins, S. (1979). Many hands make light the work: The causes and consequences of social loafing. *Journal of Personality and Social Psychology*, *37*(6), 822–832. https://doi.org/10.1037/0022-3514.37.6.822

Kerr, N. L., & Bruun, S. E. (1983). Dispensability of member effort and group motivation losses: Free-rider effects. *Journal of Personality and Social Psychology*, *44*(1), 78–94. https://doi.org/10.1037/0022-3514.44.1.78

Kollar, I., Fischer, F., & Hesse, F. W. (2006). Collaboration scripts – A conceptual analysis. *Educational Psychology Review*, *18*(2), 159–185. https://doi.org/10.1007/s10648-006-9007-2

Kovanovic, V., Gašević, D., Dawson, S., Joksimovic, S., & Baker, R. (2015). Does time-on-task estimation matter? Implications on validity of learning analytics findings. *Journal of Learning Analytics*, *2*(3), 81–110. https://doi.org/10.18608/jla.2015.23.6

Stallings, J. (1980). Allocated academic learning time revisited, or beyond time on task. *Educational Researcher*, *9*(11), 11–16. https://doi.org/10.3102/0013189X009011011

Suthers, D. D. (2006). Technology affordances for intersubjective meaning making: A research agenda for CSCL. *International Journal of Computer-Supported Collaborative Learning*, *1*(3), 315–337. https://doi.org/10.1007/s11412-006-9660-y

Wise, A. F., Knight, S., & Shum, S. B. (2021). Collaborative learning analytics. In U. Cress, C. Rosé, A. F. Wise, & J. Oshima (Eds.), *International handbook of computer-supported collaborative learning* (pp. 425–443). Springer. https://doi.org/10.1007/978-3-030-65291-3_23

Acknowledgments

This work was funded by Swiss National Science Foundation (SNSF) grant 407740_187534. We thank the instructors, teaching assistants, and students for their continued support.

International Society of the Learning Sciences

From Initiative to Impact: Modeling How Collaborative Initiative Shapes Problem-Solving in STEM Simulations

Rodrigo Hidalgo, Jina Kang
rh24@illinois.edu, jinakang@illinois.edu
University of Illinois Urbana-Champaign, USA

Abstract: This study examines how collaborative learning unfolds in HoloOrbits, a WebGL-based computer-supported collaborative learning (CSCL) environment designed to teach planetary motion. Using Epistemic Network Analysis (ENA), we analyzed how students initiated (collaborative intiative; CI) and responded to one another's contributions (uptake; UP) during group inquiry. We found most initiatives were procedural, focusing on planning, clarification, and task management, while adaptive and idea-generating initiatives occurred less frequently. Groups where students' contributions were taken up and extended by peers engaged in more sustained, collective reasoning; whereas others showed more confirmatory or fragmented discourse. These findings position collaborative initiative as a relational construct, with value that lies not in the correctness of individual contributions but in how peer responses take up and build on them to support collective sensemaking. The study underscores the importance of designing CSCL environments that scaffold both student agency and responsiveness, enabling individual contributions to become shared understanding.

Keywords: Computer-Supported Collaborative Learning; Collaborative Initiative; Uptake; Epistemic Network Analysis; STEM Education

Introduction

Three-dimensional interactive simulations can support spatial reasoning and shared problem-solving in CSCL environments, and technologies such as WebGL have made them increasingly accessible directly through web browsers without additional software (Hennig et al., 2013). Yet fostering productive group work in these settings remains a persistent challenge: students often struggle to coordinate actions, integrate multiple viewpoints, and build shared understanding—especially in simulation-based settings that impose both conceptual and interface-related cognitive demands (Moro et al., 2023; Sun et al., 2020).

In such environments, knowledge building depends as much on peer uptake as on individual contributions. This study therefore examines two interrelated constructs: Collaborative Initiative (CI) and Uptake (UP). CI captures spontaneous, unprompted contributions through which students advance group activity—such as proposing strategies, sharing divergent ideas, or inviting contributions (Chu-Carroll & Brown, 1997; Rajarathinam & Kang, 2025). UP refers to how peers respond to and build on those contributions (Barron, 2003; Chan et al., 2019). Unlike participation or engagement, which describe the extent and quality of involvement in a task (Fredricks et al., 2004), CI and UP together capture the interactional processes through which ideas were introduced, negotiated, and developed into shared understanding. In this sense, CI represents directional moves that reorient the group's epistemic or task focus—the micro-level mechanisms through which students exercise epistemic agency (Scardamalia & Bereiter, 2006; Zhang et al., 2009) and enact unscripted, self-initiated regulation beyond what structured frameworks anticipate (Järvelä & Hadwin, 2013; Chu-Carroll & Brown, 1997; Howard et al., 2017).

Examining CI alongside UP, this study investigates how their interplay shapes whether individual contributions become, or fail to become, shared understanding in collaborative inquiry. Two research questions guided our analysis:

- **RQ1.** What types of collaborative initiative behaviors emerge as students interact with a science simulation during collaborative inquiry?
- **RQ2.** How do peers respond to students' collaborative initiative moves, and how do these interaction patterns differ across groups?

Methods

This study investigated collaborative processes within HoloOrbits, a WebGL-based CSCL environment. The platform guides students through inquiry-based tasks related to planetary motion and Kepler's Laws. Its manipulable 3D interface afforded direct parameter experimentation—a design feature that, as we discuss later,

 © ISLS

shaped the forms of initiative that emerged. Eighteen undergraduate students (13 male, 5 female) with no prior physics or astronomy background were randomly assigned to six groups and completed a 50-minute, four-part inquiry sequence. Facilitators provided procedural support only; all sessions were recorded and transcribed.

Drawing on Chu-Carroll & Brown (1997) and Rajarathinam & Kang (2025), we distinguish task initiative (steering group progress through planning, proposing, or directing action) from dialogue initiative (managing communication flow, clarifying meaning, or inviting peers to contribute). Six CI sub-facets were operationalized (see Table 1). Not every utterance was coded as CI or UP; codes applied only when a contribution met each construct's definitional criteria. Student Uptake (UP) was coded using Barron's (2003) framework: Accept (A) for agreement, Discuss (D) for acknowledgment with elaboration, Reject (R) for dismissal without rationale, and Ignore (I) for absence of acknowledgment.

Table 1
Collaborative Initiative Sub-Facets with Initiative Type and Examples.

Sub-facet [Code]	Type	Description	Example utterance
Proposing Strategies & Plans [PSP]	Task	Offering structured steps or plans to address the task.	*"I guess, figure out if the positions on the ellipse are similar to that of the solar system? that can be our plan."*
Demonstrating Flexibility [DS]	Task	Modifying or adapting ideas in response to new challenges or peer input.	*"I was thinking circle before, but now that we plotted more points, it looks more like an ellipse."*
Sharing Divergent Ideas [SDI]	Task	Introducing unconventional or novel ideas to the group.	*"Another idea is to flip the axes and check if it still appears elliptical."*
Exploring the Problem Space [EPS]	Task	Testing or manipulating elements of the task to investigate possibilities.	*"Let me move this and see if it stretches."*
Seeking Help or Clarification [SHC]	Dialogue	Requesting support, explanations, or feedback when uncertain.	*"How am I supposed to know about the motion based on this…?"*
Inviting Contributions [IC]	Dialogue	Prompting peers to share their thoughts or participate more actively.	*"What do you all think, which focus should we pick?"*

We segmented all transcripts into utterances—defined as a single speaker's contribution bounded by a change in speaker or sentence completion—and coded each for CI and UP. Across six groups (~400 utterances each; ~2,400 total), CI behaviors accounted for approximately 38% of coded utterances (n = 912), distributed as: PSP (41%), EPS (24%), SHC (19%), SDI (7%), DS (5%), IC (4%). Three coders established inter-rater reliability ($\kappa > .85$ across all categories) through two iterative calibration rounds; discrepancies were resolved collaboratively before coders continued independently.

Epistemic Network Analysis (ENA) quantifies the co-occurrence of discourse codes within conversational contexts (Shaffer et al., 2016; Ruis et al., 2019; Siebert-Evenstone et al., 2017). In ENA networks, nodes represent discourse codes and edge weights (line thickness) reflect co-occurrence frequency within a sliding conversational window—thicker edges indicate codes that cluster more often. We used a four-line sliding window with the utterance as the unit of analysis. RQ1 mapped CI sub-facet co-occurrence patterns; RQ2 modeled CI and UP together to examine how initiative moves were received across groups.

Results

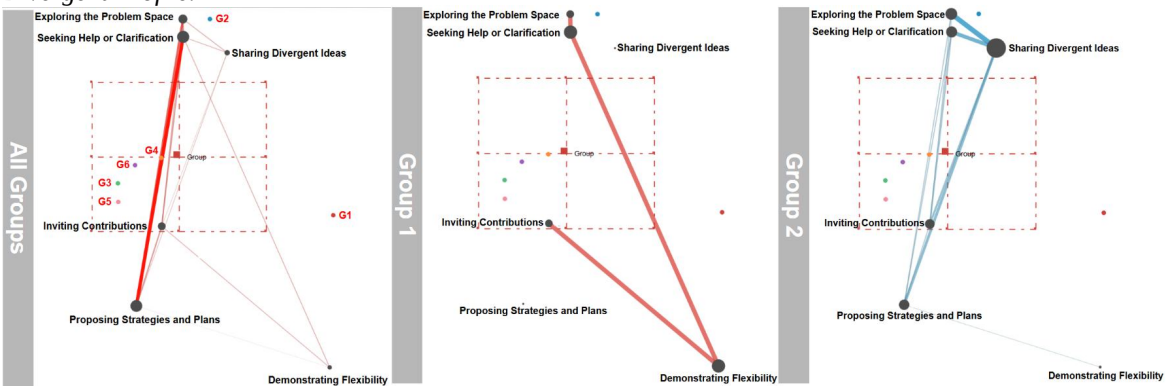
RQ1. Collaborative initiative behaviors during simulation-based inquiry

RQ1 aims to identify the types of collaborative initiative (CI) behaviors students exhibit as they engage in inquiry using the science simulation. Specifically, it explores how students take spontaneous, unprompted actions—such as proposing strategies, exploring ideas, or inviting contributions—to advance group problem-solving and knowledge building. By mapping these behaviors, RQ1 seeks to characterize the forms of initiative that drive collaboration within a CSCL environment. The aggregate ENA plot (Figure 1, left) shows that most student initiatives clustered around PSP, EPS, and SHC. This pattern reflects a predominantly procedural orientation to collaboration—consistent with prior findings that task-regulatory talk dominates simulation-based inquiry

(Mercer, 1996; Roschelle & Teasley, 1995). SDI and DS were less central, while IC occupied an intermediate bridging role.

Figure 1

Left side: Aggregate ENA Network of CI Sub-Facets Across Six Student Groups. Nodes Represent Each Sub-Facet; Edge Weight Reflects Co-Occurrence Frequency. Center: Group 1 Adaptive Profile. Right: Group 2 Divergent Profile.



Group-level comparisons reveal distinct profiles of adaptation and divergence. Group 1 (Figure 1, center) showed an *adaptive initiative* profile: DS co-occurred strongly with SHC and IC, indicating that students flexibly revised plans while actively soliciting peer involvement. Group 2 (Figure 1, right) reflects a *divergent initiative* profile, with SDI co-occurring strongly with EPS and SHC: students frequently introduced speculative framings that prompted exploration rather than task execution.

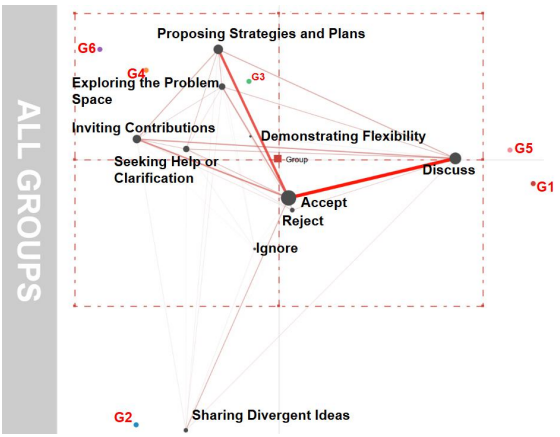
RQ2. Peer responsiveness to collaborative initiative across groups

RQ2 explores how CI moves were responded to by peers across groups, focusing on patterns of uptake and their variation. While RQ1 examined the kinds of CI moves present in student discourse, RQ2 shifts attention to how those moves functioned within group interactions, specifically, whether they were acknowledged, elaborated, or left unaddressed. The aggregate ENA network of CI and UP (Figure 2) shows strong co-occurrence between Accept (A) and Discuss (D), reflecting baseline acknowledgment patterns across all groups. Group centroid positions reveal meaningful variation: groups positioned near uptake nodes (A, D) tended to affirm contributions; groups positioned near CI sub-facets showed more generative initiative-taking.

Group 6 (Figure 3, right) displayed the most integrated CI-UP coupling: PSP co-occurred strongly with Accept (strength = .61) and Discuss, suggesting sustained cycles of proposing, acknowledging, and elaborating consistent with productive knowledge building (Chan et al., 2019; Scardamalia & Bereiter, 2006). Group 4 showed similar but weaker patterns (PSP–A strength = .32), where contributions were acknowledged but less often extended. Group 3's discourse centered on a PSP–A–D triad (PSP–A = .79; A–D = .47) with limited diversification across CI sub-facets.

Figure 2

Aggregate ENA Network of CI and UP Sub-Facets: All Six Groups

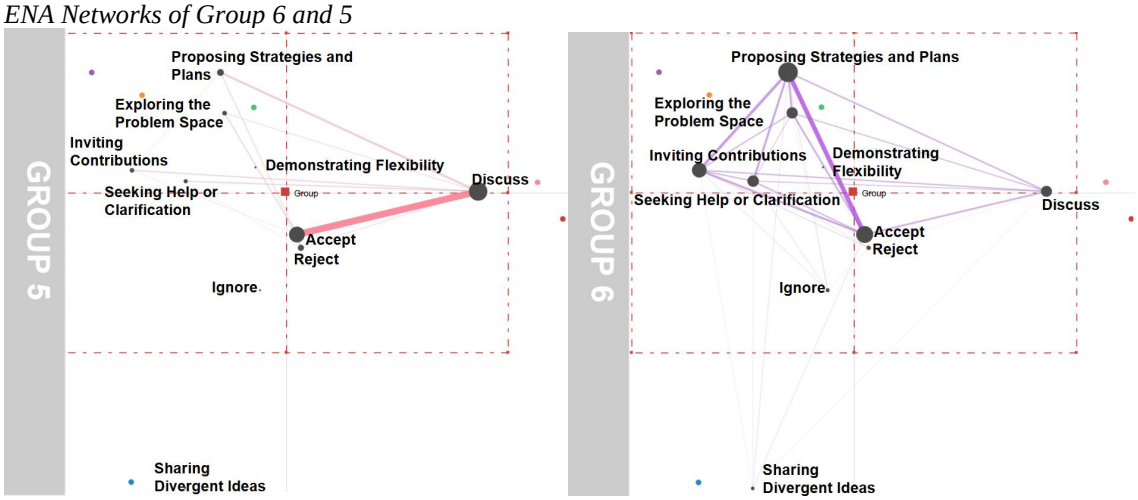


 © ISLS

Group 5 (Figure 3, left) displayed weak CI–UP integration: Accept–Discuss cycles were largely disconnected from initiative moves, suggesting surface-level collaboration without sustained uptake of peer contributions. Group 1 showed a similar pattern. Beyond these patterns, Group 2's profile was distinct: divergent ideas were accepted but rarely developed, producing a creative yet fragile dynamic in which novel contributions circulated without gaining collective traction.

Figure 3

ENA Networks of Group 6 and 5



Discussion and conclusion

This study examined how CI and UP interact to shape collaborative inquiry in HoloOrbits. Across groups, task-related initiatives (PSP, EPS, SHC) dominated, reflecting a procedural orientation to collaboration. More relational moves (IC, DS, SDI) were less frequent but consequential: groups that deployed these moves alongside responsive uptake showed richer collaborative dynamics, while procedurally focused groups often settled into confirmatory cycles.

Our findings support a relational view of CI: initiative gains significance through peer uptake. This aligns with knowledge building theory (Scardamalia & Bereiter, 2006; Zhang et al., 2009), where ideas must receive elaborative uptake to contribute to collective understanding rather than remain individual assertions. The contrast between adaptive and divergent initiative also resonates with work on epistemic agency in CSCL (Järvelä & Hadwin, 2013), where unscripted, self-initiated moves are understood as mechanisms through which students shape shared knowledge outside standard regulatory sequences. The value we attribute to CI is not correctness, but contribution to collective sense-making: an initiative becomes epistemically productive when it is acknowledged, extended, and integrated into the group's ongoing reasoning.

The platform's affordances shaped these patterns: HoloOrbits' manipulable 3D interface supported exploratory manipulation and EPS behaviors, while SDI and IC were comparatively rare. This suggests the task and interface scaffolded execution more than divergent thinking or peer invitation, highlighting opportunities for future CSCL designs to incorporate prompts that better support adaptive and divergent initiative.

As an exploratory study with six groups from a single institution, these findings are not generalizable. Future research across larger datasets and CSCL contexts can further examine how peer responsiveness shapes learning. Together, CI and UP reveal interactional mechanisms that participation-based frameworks often overlook—how student agency becomes shared understanding or dissipates in fragmented discourse.

References

Barron, B. (2003). When smart groups fail. *The Journal of the Learning Sciences*, *12*(3), 307–359. https://doi.org/10.1207/S15327809JLS1203_1

Chan, C. K. K., Yang, Y., & Wang, S. (2019). Studying uptake in collaborative learning. *Instructional Science*, *47*(2), 165–185. https://doi.org/10.1007/s11251-018-9476-9

Chu-Carroll, J., & Brown, M. K. (1997). Tracking initiative in collaborative dialogue interactions. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics* (pp. 262–270). Association for Computational Linguistics.

 © ISLS

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. Review of Educational Research, 74(1), 59–109. https://doi.org/10.3102/00346543074001059

Hennig, M., Gaspers, D., & Mertsching, B. (2013). Interactive WebGL-based 3D visualizations for situated mathematics teaching. In *Proceedings of the 12th International Conference on Information Technology Based Higher Education and Training (ITHET)* (pp. 1–6). IEEE. https://doi.org/10.1109/ITHET.2013.6671038

Howard, C., Di Eugenio, B., Jordan, P., & Katz, S. (2017). Exploring initiative as a signal of knowledge co-construction during collaborative problem solving. *Cognitive Science, 41*(6), 1422–1449. https://doi.org/10.1111/cogs.12400

Järvelä, S., & Hadwin, A. F. (2013). New frontiers: Regulating learning in CSCL. *Educational Psychologist, 48*(1), 25–39. https://doi.org/10.1080/00461520.2012.748006

Mercer, N. (1996). The quality of talk in children's collaborative activity in the classroom. *Learning and Instruction*, *6*(4), 359–377. https://doi.org/10.1016/S0959-4752(96)00021-7

Moro, C., Alrutz, M., Zhang, M., & Gregory, B. (2023). Immersive technology for learning: Cognitive load and usability. *Educational Technology Research and Development, 71*(2), 419–439. https://doi.org/10.1007/s11423-023-10148-6

Rajarathinam, V., & Kang, Y. (2025). Defining collaborative initiative: A framework for analyzing unprompted contributions in CSCL environments. In *Proceedings of the International Conference of the Learning Sciences (ICLS 2025)*. https://doi.org/10.5281/zenodo.14966499

Roschelle, J., & Teasley, S. D. (1995). The construction of shared knowledge in collaborative problem solving. In C. E. O'Malley (Ed.), *Computer-supported collaborative learning* (pp. 69–97). Springer. https://doi.org/10.1007/978-3-642-85098-1_5

Ruis, A. R., Siebert-Evenstone, A. L., & Shaffer, D. W. (2019). Finding common ground: A method for quantifying the semantics of conversation. In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge* (pp. 123–139). ACM. https://doi.org/10.18608/jla.2019.5416

Scardamalia, M., & Bereiter, C. (2006). Knowledge building: Theory, pedagogy, and technology. In K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 97–115). Cambridge University Press.

Shaffer, D. W., Collier, W., & Ruis, A. R. (2016). A tutorial on epistemic network analysis: Analyzing the structure of connections in cognitive, social, and interaction data. *Journal of Learning Analytics, 3*(3), 9–45. https://doi.org/10.18608/jla.2016.33.3

Siebert-Evenstone, A. L., Arastoopour, G., Shaffer, D. W., & Ruis, A. R. (2017). In search of conversational grain size: Modeling semantic structure using Epistemic Network Analysis. *Journal of Learning Analytics, 4*(3), 123–139. https://doi.org/10.18608/jla.2017.43.7

Sun, C., Wang, S., & Chan, C. K. K. (2020). A framework for assessing collaborative problem-solving processes and outcomes. *Educational Psychology Review, 32*, 939–961. https://doi.org/10.1007/s10648-020-09522-0

Zhang, J., Scardamalia, M., Reeve, R., & Messina, R. (2009). Designs for collective cognitive responsibility in knowledge-building communities. *Journal of the Learning Sciences, 18*(1), 7–44. https://doi.org/10.1080/10508400802581676

Acknowledgement

This work was partially supported by the National Science Foundation under Grant DRL-2229612. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

International Society of the Learning Sciences

A Transpositional Grammar Cyber-Social Study of Student–AI Relationships

Christopher Hughes, University of Illinois, chughes7@illinis.edu
Akash Saini, University of Illinois, aksaini2@illinois.edu

Abstract: As AI-mediated feedback becomes embedded in higher education, understanding how learners construe relationships with AI and human reviewers becomes increasingly more important. This pilot applies Transpositional Grammar (TG) to student prompts, AI-generated images, and reflections in a cyber-social learning environment. Graduate students (n = 7; P300–P306) at a Midwestern U.S. research university contributed coursework artifacts coded collaboratively by a human researcher and GPT-5, yielding Cohen's $\kappa = 0.761$. TG analysis identifies five recurrent transpositions—expansion, visualization, reinterpretation, materialization, and affective deepening—and surfaces *relational accountability* as a new dimension of learner expectations. The full study (n = 35; P300–P334) will add a second human coder to triangulate interrater reliability and inform the design of affect-sensitive human–AI feedback ecosystems in graduate instruction.

Introduction

AI-supported feedback now shapes core practices in higher education, yet we know comparatively little about how learners represent their relationships with AI tools and human reviewers across modalities. In cyber-social learning environments, feedback is simultaneously interpersonal and computational, blending peer interaction, AI assistance, and platform analytics into a coupled socio-technical process (e.g., CGScholar; Cope & Kalantzis, 2023). This hybrid ecology warrants multimodal analysis attentive to how meanings travel between prompts, image, and reflection (Zapata et al., 2025).

Transpositional Grammar (TG; Cope & Kalantzis, 2020; Kalantzis & Cope, 2020) offers a principled lens for tracing such cross-modal shifts. By modeling meaning functions, Reference, Agency, Structure, Context, and Interest, and their transpositions, TG equips us to examine how students' depictions of AI–human relationships migrate from prompts to generated visuals to reflective commentary. Guided by TG, our analysis attends to five transpositions, expansion, visualization, reinterpretation, materialization, and affective deepening, and advances the construct of *relational accountability* to characterize learners' expectations of reciprocity, transparency, and responsibility in human–AI feedback.

Methodologically, the study mirrors its subject matter: researchers adopt a cyber-social approach by engaging GPT-5 as a co-researcher alongside a human coder, operationalizing distributed cognition within the analytic workflow. The pilot draws on graduate coursework artifacts from Spring 2025 at a Midwestern U.S. research university (n = 7; P300–P306), with plans to extend to the full corpus (n = 35; P300–P334) and incorporate an additional human coder to triangulate interrater reliability. This design enables us to link representational patterns to the affordances and frictions of AI-mediated feedback while maintaining methodological transparency. We address the following research questions:

RQ1: What do student prompts reveal about learners' representations of AI–human feedback relationships?
RQ2: What do AI-generated images reveal about AI's interpretation of its relationship with students?
RQ3: What do student reflections reveal about learners' interpretations of AI–human relationships?

By situating multimodal learner representations within a TG framework and a cyber-social method, the study aims to clarify how meanings about human–AI feedback is constructed and transformed across modalities, informing the design of more authentic and affect-sensitive feedback ecosystems in graduate education.

Literature review

Learner perceptions in cyber-social contexts

Across cyber-social learning contexts, learners hold nuanced and often contradictory attitudes toward AI and human feedback. Hughes et al. (2024) found that graduate students using CGScholar blended skepticism with openness toward AI-generated reviews, suggesting that acceptance of AI feedback is neither straightforward nor uniform. Zapata et al. (2024a) extended this finding, showing that while human reviews are perceived as more authentic, students nonetheless prefer a future hybrid model — revealing a tension between relational trust and practical utility. Zapata et al. (2024b) demonstrated through *activity theory* that calibrated GenAI feedback can be

 © ISLS

perceived as a community member within feedback activity systems, and Tzirides et al. (2024) showed that combining human and AI tools enhances AI literacy and critical assessment capacity. Taken together, these studies show that learners evaluate AI feedback through relational expectations, not accuracy alone. However, prior work has not examined how those expectations are constructed and transformed across modalities, or what learners mean when they hold AI accountable for affective and relational norms. Our study addresses this gap.

AI as qualitative co-researcher

Emerging work validates AI as a coding partner. Cho (2025) co-coded classroom dialogue with ChatGPT, reporting $\kappa = 0.56$. Ni et al. (2024) introduced QualiGPT for inductive/deductive coding of online community posts, achieving $\kappa = 0.57$ (inductive) and Fleiss' $\kappa = 0.46$ (deductive). Xie et al. (2024) applied chain-of-thought prompting with GPT-4, achieving substantial agreement ($\kappa \geq 0.60$) across eight codes and excellent agreement ($\kappa \geq 0.79$) for three codes. These benchmarks guided our human–AI collaborative coding design and establish that $\kappa = 0.76$ represents strong inter-rater reliability for complex multimodal qualitative analysis.

Transpositional grammar in multimodal research

TG articulates five interrelated functions of meaning across modes (Cope & Kalantzis, 2020; Kalantzis & Cope, 2020). Transpositional Grammar offers three analytical advantages over related frameworks that make it particularly well-suited to this study's aims. First, unlike Kress and van Leeuwen's (2001) *social semiotic multimodal analysis*, which focuses on the relatively stable grammatical properties of specific modes, TG conceptualizes meaning forms as inherently migratory — constantly transposing across text, image, space, body, sound, and speech rather than remaining fixed within any single mode (Hughes, 2025). Second, where O'Halloran's (2008*) systematic functional-multimodal discourse analysis* foregrounds the co-occurrence of modes within a single text, TG traces the *transformative processes* that occur when meanings shift across modalities — precisely what this study requires in moving from prompts (P) to images (I) to reflections (R). Third, while discourse analysis (e.g., Halliday & Hasan, 1980) remains primarily oriented toward language, TG's five meaning functions — Reference, Agency, Structure, Context, and Interest — operate simultaneously across all modalities, enabling a unified analytical language for multimodal, cyber-social data (Cope & Kalantzis, 2020; Kalantzis & Cope, 2020). While prior applications analyze text and image in educational settings, none examine student–AI feedback relationships. TG's focus on how meaning shifts through transpositions (P→I→R) makes it ideal for capturing dynamic relational constructions in cyber-social feedback contexts.

Methods

Context, sample, and ethics

This pilot analyzes a 20% sample (n = 7, P300–P306) from a graduate cyber-social learning course at a large U.S. public research university (Spring 2025); the full study expands to the complete corpus (n = 35, P300–P334). Participants were graduate students and working professionals in education technology and learning design–related fields. Generative AI was integrated throughout the course: students drafted ≥3,000-word essays, received rubric-aligned RAG-enhanced AI feedback, and completed blind peer review in CGScholar. Afterward, students generated AI images depicting experiences of AI versus human feedback and wrote accompanying explanations; these images and reflections are analyzed here. The study received IRB approval (IRB24-1772). Participants consented via a voluntary pre-course Qualtrics survey, could opt out without penalty, and all data were de-identified prior to analysis; the research team held no instructional or grading responsibilities.

Data layers

1. Prompts (Layer 1): Students wrote textual requests for AI to visually depict their perceived AI–human reviewer relationships.
2. Images (Layer 2): Students were free to use any AI image-generation tool (e.g., ChatGPT, Leonardo, Gemini). This reflects authentic cyber-social practice in which learners mix tools rather than operate within a single platform. This cross-platform variability is treated as a boundary condition but does not compromise the TG framework, which focuses on meaning functions rather than stylistic fidelity.
3. Reflections (Layer 3): Student-written evaluations of the AI-generated images.

Transpositional grammar framework

TG analyzes five functions: Reference (entities/properties), Agency (actions/status), Structure (relations/grouping), Context (setting/modality), and Interest (affect/stance). Each data layer was coded

for these functions. Transposition tags logged meaning shifts: P→I (prompt to image), I→R (image to reflection), and P→R (prompt to reflection).

Collaborative coding and Inter-Rater Reliability

R1 (human researcher) and R2 (GPT-5) independently applied an iteratively refined TG codebook across all pilot layers, followed by adjudication. Refinements clarified evidence rules (e.g., assigning status only when explicit control/deference cues are present), distinguished Context versus Structure indicators, and standardized use of multimodal mix, symbolism/metaphor, and transposition tags. To reduce disagreements driven by paraphrase rather than concept, inter-rater reliability was calculated after semantic normalization: synonymous labels were collapsed, "absence" labels were standardized, prompt/reflective tokens were treated as matches when overlap met a Jaccard threshold of ≥ 0.40, and image entity sets were treated as equivalent when one set was a subset of the other. This approach aligns with multimodal qualitative coding where constructs can be expressed in multiple valid phrasings and yielded Cohen's $\kappa = 0.76$. Future work adds a second human coder (R3) for triangulated IRR on the full corpus (P300–P334).

Findings

Inter-Rater Reliability

Human–AI collaborative coding achieved substantial agreement (Cohen's $\kappa = 0.76$), supporting the robustness of the co-researcher approach for complex multimodal data.

Prompts

Student prompts consistently frame AI and human reviewers as comparable yet distinct agents. Three themes emerge from the prompt data: (1) Role Delineation: students explicitly partition AI and human reviewers into distinct but comparable roles; (2) Legitimacy Negotiation: prompts reveal students working through the credibility of AI as a feedback provider; and (3) Structural Framing: students consistently invoke contrast language ("vs.," "compare," "split-screen") to organize their representations. Prompts such as "Show a split-screen with an AI reviewer on the left and a human reviewer on the right, each highlighting strengths and weaknesses of my writing" invoke TG's Reference, Agency, and Structure functions by naming entities (AI, human reviewer), actions (review, highlight), and explicit contrasts ("vs," "compare"). These framing strategies reveal students' need to delineate roles while negotiating the legitimacy of AI feedback.

Images

AI-generated images literalize the prompts' contrasts but often transcend them by depicting collaborative configurations. Two themes emerge from the image data: (1) Literalization: AI outputs faithfully render the contrasts requested in prompts; and (2) Collaborative Transcendence: images frequently exceed the prompt's contrast logic by depicting AI and human reviewers in configurations of co-presence and equal partnership. One representative AI-generated image from P305 positions a humanoid robot and two human peers equally around a book, each pointing or reading, a triadic scene of shared inquiry.

Reflections

In reflections, students validate the AI's faithful mapping of prompt semantics but introduce critical relational critiques. Two themes emerge from the reflections: (1) Fidelity Validation: students affirm the AI's accurate mapping of their prompt semantics; and (2) Relational Critique: students introduce critical evaluations of emotional expressiveness, identity representation, and reciprocal agency that were absent from the original prompts. One student wrote, "I appreciate the depiction of all reviewers collaborating, but the AI looks too robotic; it should show more warmth." These reflections engage TG's Interest and Context functions through evaluative stance and experiential references, highlighting expectations for emotional expressiveness, accurate identity representation, and reciprocal agency.

Relational dynamics and novel insight

Table 1 summarizes how each TG function transforms across Prompt (P) → Image (I) → Reflection (R). Readers are directed particularly to the 'Shift' column, which captures the primary novel insight of this analysis: five predictable, named transpositions: *expansion*, *visualization*, *reinterpretation*, *materialization*, and *affective deepening*, that characterize how relational meaning migrates across semiotic layers in cyber-social feedback contexts.

Table 1
Transpositional Grammar Analysis Across Semiotic Layers

TG Function	Prompt (P)	Image (I)	Reflection (R)	Shift (P→I→R)
Reference (Entities, Properties)	"AI," "human," "reviewer"	Robots, people, tools visually	Entities with platform context	Expansion: generic terms → visual detail → platform-specific names
Agency (Actions, Status)	"review," "collaborate"	Spatial co-presence or split panels	Iterative self-positioning	Visualization: verbal actions → spatial arrangements → temporal sequences
Structure (Contrast, Grouping)	"vs," "compare"	Side-by-side contrasts, parallel tools	Complementarity emphasis	Reinterpretation: contrast request → visual split → both-and complementarity
Context (Setting, Modality)	Classroom, module references	Flat/photoreal styles; platform labels	Past experience, platform names	Materialization: implied setting → concrete visual style → experiential framing
Interest (Affect, Stance)	"helpful," "authentic," "precise"	Facial expressions: smiling vs. neutral	Affective critique ("AI should smile")	Affective deepening: adjectives → facial expressions → critical stance on authenticity

The analysis in Table 1 reveals a progression from contrast to complementarity to accountability. Prompts (P) explicitly request comparisons, positioning AI as "other" yet comparable to human reviewers, while images (I) literalize contrasts but frequently depict collaborative arrangements that signal openness to AI as a peer-like partner. Reflections (R) affirm visual accuracy while introducing critical affective and contextual dimensions, holding AI accountable for emotional expressiveness, identity representation, and reciprocal agency. This gives rise to relational accountability as a construct: learners expect AI feedback systems to honor human-like relational norms to foster trust and engagement.

Discussion

This pilot study offers two contributions, aligned with the three research questions. Prompts reveal learners' role delineation, legitimacy negotiation, and structural framing, positioning AI as a comparable yet distinct feedback agent. Images literalize prompt contrasts but often transcend them through collaborative configurations, suggesting partnership-oriented relational defaults. Reflections introduce relational critique, holding AI accountable for affective resonance, identity alignment, and reciprocal agency—what we term *relational accountability* in this context. These findings underscore relational accountability as a central design concern for computer-supported collaborative learning systems that integrate AI as a feedback partner.

Limitations and future work: First, the pilot sample (n = 7, P300–P306) comes from one graduate course at a Midwestern U.S. research university, limiting generalizability across contexts. Second, the small participant group means the five transpositional shifts and the relational accountability construct should be treated as emergent. Third, while the AI co-researcher model yielded $\kappa = 0.76$, it currently relies on a single human coder; the full study (n = 35, P300–P334) adds a second human coder (R3) to enable triangulated IRR. Fourth, multiple image-generation tools introduce cross-platform variability in how Agency is rendered; the full study will examine function-specific transpositions in greater depth and test TG-informed feedback tool designs in experimental settings.

Conclusion

By applying transpositional grammar to a cyber-social phenomenon using AI co-researcher methods, this pilot uncovers predictable relational transpositions across modalities and introduces *relational accountability* as a critical dimension of student–AI feedback interactions. Full-scale analysis and design experimentation will follow, advancing both multimodal learning sciences and human–AI interaction design.

 © ISLS

References

Cho, H. (2025). Co-coding classroom dialogue: A single researcher case study of ChatGPT-assisted analysis in science education. *Journal of Computer Assisted Learning*, 41. https://doi.org/10.1111/jcal.70089

Cope, B., & Kalantzis, M. (2020). *Making sense: Reference, agency, and structure in a grammar of multimodal meaning*. Cambridge University Press. https://doi.org/10.1017/9781316459645

Cope, B., & Kalantzis, M. (2023). Creating a different kind of learning management system: The CGScholar experiment. In *Promoting next-generation learning management systems* (pp. 1–28). Routledge.

Halliday, M. A. K., & Hasan, R. (1980). *Cohesion in English.* Longman.

Hughes, C. (2025). Meaning Particles and Waves in MOOC Video Lectures: A transpositional grammar guided observational analysis. *Computers & Education, 236*, 105308.

Hughes, C., Saini, A. K., Tzirides, A.O. (2024). A Quantitative Investigation of Graduate Student Perceptions of Human-Generated and AI-Generated Reviews in a Cyber-Social Learning Platform. In *Trust and Inclusion in AI-mediated Education: Where Human Learning Meets Learning Machines* (pp. 213-234). Cham: Springer Nature Switzerland.

Kalantzis, M., & Cope, B. (2020). *Adding sense: Context and interest in a grammar of multimodal meaning*. Cambridge University Press. https://doi.org/10.1017/9781108862059

Kress, G., & van Leeuwen, T. (2001). *Multimodal discourse: The modes and media of contemporary communication.* Arnold.

Ni, J., Ni, C., Zhu, J., Huang, J., Carroll, J. M., & Cai, J. (2024). Exploring inductive and deductive qualitative coding with AI: Investigating inter-rater reliability between large language model and human coders. In *Proceedings of the 2024 International Conference on Human-Computer Interaction*.

O'Halloran, K. L. (2008). Systemic functional–multimodal discourse analysis (SF-MDA). *In L. Unsworth (Ed.), Multimodal semiotics* (pp. 1–30). Continuum.

Tzirides, A. O., Zapata, G. C., Kastania, N. P., Saini, A. K., Castro, V., Ismael, S. A. R., You, Y., Afonso dos Santos, T., Searsmith, D., O'Brien, C., Cope, B., & Kalantzis, M. (2024). Combining human and artificial intelligence for enhanced AI literacy in higher education. *Computers and Education Open*, 6, 100184. https://doi.org/10.1016/j.caoe.2024.100184

Xie, H., Abuhamad, S., Liu, T., Happ, G., & Zhang, W. (2024). Scalable qualitative coding with LLMs: Chain-of-thought reasoning matches human performance in some hermeneutic tasks. *arXiv preprint arXiv:2401.15170*. https://arxiv.org/abs/2401.15170

Zapata, G. C., Saini, A. K., Cope, B., Kalantzis, M., & Searsmith, D. (2024a). Peer-reviews and AI feedback compared: University students' preferences. *EdArXiv*. https://doi.org/10.35542/osf.io/preprint

Zapata, G. C., Saini, A. K., Tzirides, A. O., Cope, B., & Kalantzis, M. (2024b). The role of AI feedback in university students' learning experiences: An exploration grounded in Activity Theory. *Ubiquitous Learning: An International Journal, 18*(2), 1–30.

Zapata, G. C., Cope, B., Kalantzis, M., Tzirides, A. O., Saini, A. K., Searsmith, D., Whiting, J., Kastania, N. P., Castro, V., Kourkoulou, T., Jones, J. W., & Abrantes da Silva, R. (2025). AI and peer reviews in higher education: Students' multimodal views on benefits, differences and limitations. *Technology, Pedagogy and Education*. https://doi.org/10.1080/1475939X.2025.2480807

International Society of the Learning Sciences

Who Gets to Know? Children's Epistemic Agency in Science eBook Reading with AI

Xuechen Liu, University of Michigan, liuxuech@umich.edu
Ying Xu, Harvard University, yingxu@g.harvard.edu
Jon M. Wargo, University of Michigan, wargojon@umich.edu

Abstract: Children's ability to act as epistemic agents not only support their understanding of concepts but also for shaping how they engage with ideas and participate in knowledge construction within their learning environment. The advancement in artificial intelligence (AI) has provided opportunities for interactive science learning, such as serving as a learning companion in dynamic and dialogic reading. While prior research has examined how children enact epistemic agency in human-to-human interaction, it is unclear about whether, and if so how, children enact epistemic agency when interacting with AI, a non-human entity. This paper draws on video recordings of two individual children, each reading a science eBook with an AI agent. Using an interaction analysis, we examine how, through both language and embodied action, these two children claimed, evaluated, and negotiated knowledge, positioning themselves as epistemic agents in interaction with the AI agent.

Introduction

Epistemic agency emphasizes that learners have the capacity to contribute to knowledge building through actions like asking questions, evaluating ideas, and taking responsibility for what and how they know (Scardamalia, 2002; Scardamalia & Bereiter, 1991). This ability is important in science learning, where children move beyond receiving information to constructing and refining understanding (Damșa et al., 2010; Stroupe, 2014). Prior work has examined how teachers support epistemic agency by positioning students as knowledge builders (Bautista & Davis, 2024; Stroupe, 2014), and how students enact and negotiate this agency in interaction with others (Baze et al, 2025; Jung et al., 2023).

Recent advances in AI have provided opportunities for dialogic reading, where AI can respond in real time based on the context. Carefully designed AI systems can support children's science learning by prompting them to construct responses, ask questions, and engage with feedback and scaffolding. Because epistemic agency is deeply shaped by the social and interactional context (Calabrese Barton & Tan, 2010), the nature of the learning partner—whether human or non-human—may influence how children position themselves as knowers. It is therefore important to examine whether, and if so, how children enact epistemic agency, and what this reveals about how knowledge, authority, and participation are negotiated and constructed in child–AI interactions.

To explore these questions, we analyzed video recordings of two children reading a science eBook with an AI agent using an interactional analysis approach (Jordan & Henderson, 1995). This analysis shows how children, through language and embodied action, claimed, evaluated, and negotiated knowledge—positioning themselves as epistemic agents in interaction with the AI. It also revealed how they managed their epistemic stance differently with the AI and a human researcher, showing how they wanted to be perceived as knowers, and by whom.

Epistemic agency in learning

Epistemic agency refers to learners' responsibility and ability to actively participate in knowledge-building (Stroupe, 2014; Scardamalia, 2002; Scardamalia & Bereiter, 1991). Rather than passively receiving information, learners act as epistemic agents who pose questions, evaluate explanations, and negotiate ideas. In science learning, this may involve challenging received explanations, generating alternative hypotheses, or reasoning through what they have heard. Building on this, we conceptualize the *enactment of epistemic agency* as a relational process in which children evaluate received information and negotiate and justify knowledge claims through language and embodied actions in response to the AI's questions and feedback.

Understanding children's epistemic agency through co-operative action

This study draws on Goodwin's theory of *co-operative action* (Goodwin, 2013, 2018) to examine how children enact epistemic agency in interaction with an AI agent. This notion is that, in conversation settings, action is not produced in isolation but participants collaboratively construct it using multiple semiotic resources, such as talk, gaze, gesture, and body orientation, while responding to each other (Goodwin 2000, 2012). Participants "build

 © ISLS

action by laminating different kinds of meaning making resources together" (Goodwin, 2013, p. 16), allowing us to trace how ways of knowing and acting emerge through interaction.

A key concept in co-operative action is "substrate," which refers to semiotic or material resources produced in prior turns that can be reused and transformed by subsequent participants (Goodwin, 2013). As participants build on prior actions, elements are both preserved and modified, linking new contributions to earlier ones. In this process, Goodwin mentioned two things typically happen: first, the previous substrate is modified or transformed as it is reused; second, elements of the original action are preserved in a recognizable way, linking the new act to the earlier one. This process of reusing, modifying, and preserving what came before could show how participants collaboratively build meaning over time. For instance, in child-AI conversation, the AI agent's earlier feedback can become the resource or substrate for the child's response (through language or embodied actions), making visible how this action emerged through co-operative action.

Goodwin's theory of cooperative action provides a lens for us to examine the ways children express their epistemic authority and agency with AI, as well as how they want to be perceived by the technology "AI agent". Conceptually, it allows us to see not only where a substrate first emerges in interaction, but also how it is preserved and transformed throughout the subsequent actions. In the context of this study, we can trace how the AI's prompts, feedback, or evaluations become substrates for children's responses, and how children take up or resist these substrates to claim or negotiate epistemic authority moment by moment.

Methods

This paper draws on data from the first week of our pilot study conducted prior to the launch of our large-scale experimental project on children's science learning with an AI-supported eBook. The pilot period was designed to refine study procedures and test the functionality of the eBook system. Seven children (aged 6-8) participated in individual reading sessions with the AI agent. During each session, children read the same science eBook about frogs and interacted with the AI agent, which asked questions, provided feedback, and prompted discussion throughout the reading. Children were instructed to read the eBook and respond to the AI's questions as they progressed. The two children presented here were purposely selected because their interactions highlighted the dynamics of claiming authority as knowers to AI and acted as epistemic agents. Therefore, they serve as two "telling" cases (Mitchell, 1984, p. 239), defined as instances in which particular circumstances make broader theoretical relationships visible, for examining how children enact their epistemic agency and how they want to be perceived by the AI while reading a science eBook with AI.

Data generated included video recordings of each child's individual eBook reading session with the AI agent, along with screen recordings of dialogic interaction between child and AI. Across all seven participants, recordings ranged from 16–23 minutes per session, totaling 112 minutes. These data provide a multimodal account of how children evaluate information, express understanding, and enact epistemic agency with the AI.

To understand each child's modes, bodily orientations, and interactions with the environment in relation to the AI agent's questions and responses, we followed Erickson's (2006) whole-to-part analytical procedure. The first author first watched all seven video recordings from start to finish and noted each moment when verbal and nonverbal actions occurred in a Google Doc. Shen then rewatched each of those moments in greater detail—from the beginning to the end of the relevant eBook page—when both verbal and nonverbal actions were present. For example, if a noted moment occurred when the child and the AI agent were discussing the purpose of frogs' voices, she rewatched the entire episode beginning from the start of that page to just before moving on. During this process, she identified what Jordan and Henderson (1995) referred to as "hot spots" or "shapes of events," signaling shifts in epistemic authority and agency between the child and the AI agent.

Using these selected moments as focal cases, the first author transcribed them using both Jefferson's conversation analysis conventions (2004) and Mondada's (2018) multimodal annotations. Jefferson's conversational analysis captures the temporal organization of the talk, while Mondada's framework incorporates bodily orientations and gaze shifts with the AI agent. Using both approaches allowed us to trace how epistemic authority was negotiated not only through speech but also through the embodiment and environment. Grounded in Goodwin's concept of substrates (2013), our analytic lens focuses on how AI agent's prior talk (i.e. evaluation to the child's response, re-prompt, and conclusion) become resources for subsequent turns, shaping how children position themselves epistemically and negotiate epistemic authority and meaning with the AI agent.

Findings

Presented as two "telling cases," we illustrate how children positioned themselves as knowers when interacting with the AI reading companion. In Case 1, the child asserted his epistemic authority by correcting the AI's failure to recognize the intent behind his earlier contribution, expressing dissatisfaction, and reclaiming authority. In Case 2, the child enacted epistemic agency by evaluating the AI's scaffolding utterances and claiming knowledge came

from herself rather than the AI. Across both cases, we see children are not simply responding to the AI but actively *calibrating* their relationship with their AI partner. These "telling cases" invite us to consider how children enact a sense of epistemic authority, shaping how they want to be perceived by AI.

Case 1: Expressing epistemic authority

Reading informational texts such as science books supports children's ability to move from everyday ideas toward more systematic reasoning about the physical and biological world during the elementary years. Pedagogical questioning further enhances comprehension by encouraging children to express their thinking and reflect on or revise their understanding through feedback (Carey, 2000; Vosniadou, 1994). From sharing his ideas about frogs' voices to humorously correcting the AI's feedback, the focal child in Case 1 illustrates how he sustained and defended his epistemic authority in dialogic interaction with the AI reading partner.

As evidenced in the transcript (see Extract 1a in the supplementary materials; the glossary of transcription symbols and conventions appears in the supplementary materials), the AI agent was narrating to the child (line 1) and posed a pedagogical question (line 3) to invite the child to think about the purposes of frogs' voices. While listening to the question, he moved his hand toward the "press to talk" button to get ready to respond. Immediately after the AI's question ended, he pressed the button (line 5) and gave his answer without delay or referring back to the text. He confidently shared his first response that frogs use their voices for mating, stretching the word *"mating"*. Then, the AI agent prompted the child to think about another reason frogs use their voice when they are scared (line 7). Like in the previous turn, the child moved his hand toward the "press to talk" button to get ready. Right after the question was narrated, he immediately pressed the button and gave extra force to *"scream"* (line 9) in his answer. His hand movement and ways of speaking signaled readiness and confidence in his epistemic state about this science concept, showing that he positioned himself as already knowing the answers. His answers were further confirmed by the AI's evaluation in the later turn.

However, the AI agent did not think the child had fully addressed the answer, so it prompted the child to think about another purpose (line 10). Unlike in the previous turn, this time, he was not ready to answer this question. The child paused for seven seconds while referring back to the text in search of an answer (line 11). He then responded, *"I don't really know if they use it for any other instances"* (line 12). If we just look at this textual information in this sentence alone, the phrasing shows this child's uncertainty, which is also how AI interpreted this response, as the AI agent acknowledged this uncertainty with "That's okay!". However, when combined with the child's embodied actions and ways of delivery (quick tempo, playful tone, and repeated head shakes), the child's response is not just simply expressing and signaling lack of knowledge. His response carried confidence that no additional answers existed. Based on what he knew, he thinks that there are no more other usages of frogs' voices. Through all these cues, the child's uncertainty is not as a sign of not knowing but as an epistemic stance that enabled him to sustain his role as a knower, even as the AI treated him as an unknower.

Closer to the end of the conversation on this page (see Extract 1b in the supplementary materials), the child reclaimed his authority as a knower through both facial expression, tone and language when he realized that the AI had misevaluated his second response about purposes of frogs' voices or had failed to recognize the underlying meaning the child intended. Because the system had reached the maximum number of turns, we predefined, after acknowledging the child's uncertainty, the AI proceeded to provide a fuller explanation of the correct answer, ending with a closing prompt asking whether the child had further questions (line 13). When the AI was delivering feedback, the child was showing visible frustration, furrowing his brows and silently mouthing words (line 14, fig.1-2), as if *pushing back against* this situation that the AI's interpretation failed to register the nuance in what he meant by frogs "scream" when frightened.

In other words, the child may have wondered why the AI asked for another reason for frogs' voices when, from his perspective, his earlier answers "scream" already covered what the AI later explained "to communicate" (line 13). His embodied response thus signaled frustration with how the AI evaluated and interpreted his contribution. Once the AI finished speaking, the child *pushed back again*, saying with emphasis and laughter: *"I already told you that they use their voices to communicate, ha ha ha"* (line 15, fig.3). Although his earlier response ("scream") did not explicitly mean "communicate," he framed it as equivalent to the AI's explanation of the correct answer. He laughed at AI and resisted the AI's rigid interpretation of correctness and reasserted himself as the one who knew, placing strong emphasis on the word *told*.

Case 2: Claiming authority as a knower to the AI while displaying unknowing to a human researcher

In this case (see Extract 2 in the supplementary materials), the focal child was reading a page about frogs' eyes and discussing how they help frogs see food without moving. She positioned herself as the knowledgeable one in relation to the AI by drawing on her prior understanding to explain how the eyes work, yet simultaneously

 © ISLS

displayed unknowing toward the human researcher. This invites us to think about how children may express their epistemic stance depending on how they want to be viewed by the one they are talking to.

When the AI invited the child to imagine what allows frogs to do when they can see all directions (line 1), the child paused for a moment (line 3) and responded "*I have no ↑ idea what you just said because I didn't understand it*" (line 4). Rather than simply expressing uncertainty, she shifted responsibility to the clarity of the AI's utterance. In other words, instead of passively accepting the role of an "unknower," this child demonstrated her epistemic agency by first evaluating the comprehensibility of the AI's utterance, then assessing her own understanding, and justifying her lack of ability to answer on the grounds of the AI's lack of clarity.

To give the child a hint, the AI invited the child to imagine being in the frog's situation and to think about how being able to see in all directions might help with finding food and staying safe (line 5). The child paused to think for about 14 seconds (line 7). During the first half of this time, she gazed at the chat box, reread the question, and then looked to the side toward the researcher (fig. 4), as if wanting to be recognized by the human researcher that she did not know this answer. After realizing that the researcher did not respond, she turned her gaze away and continued thinking. Since the AI did not hear anything from her, it redirected her attention back to the conversation by repeating the question (line 9).

This child used her gestures (fig. 5), facial expressions, and language to demonstrate how frogs catch food with their tongues, closing with: "*He never said some of that, but still I know that, so yeah"* (line 10). Here, she claimed her authority as a knower by distinguishing her own knowledge from the AI's, making clear that it came from her prior knowledge rather than from the AI's teaching. Her explicit statement about the source of her knowledge reinforced her epistemic agency by using confident language and embodied expression (fig.6-7) to assert that the knowledge was hers, not something learned from the AI.

Summary and discussion

The goal of this study was to examine how children enact and negotiate epistemic agency during science eBook reading with an AI partner. Rather than treating children's responses as isolated, we approached them as socially situated within an unfolding child–AI interaction, where knowledge and authority—who knows, whose knowledge counts, and how authority is displayed—are negotiated turn by turn. Although the AI was designed to scaffold understanding, children did not participate passively; they evaluated, reflected on, and at times resisted the AI's explanations, revealing a desire to be recognized as knowledgeable participants.

The two cases illustrate distinct ways children asserted epistemic agency. In one case, the child resisted the AI's explanation by asserting that his prior response was correct, signaling his expectation to be recognized as a knower. In the other, the child shifted her epistemic stance across partners—orienting uncertainty toward the human researcher while holding the AI accountable and reasserting knowledge in interaction. These moments show that epistemic authority in child–AI interaction is relational and dynamically reconstructed.

Through the lens of *co-operative action* (Goodwin, 2018), our analysis further shows that epistemic authority is enacted through multiple semiotic resources, including language, gaze, and embodied action. Children coordinated these resources to build on or challenge the AI's contributions, demonstrating how they work to sustain recognition as knowledgeable participants. Attending to these micro-level interactions offers a more nuanced account of learning with AI, highlighting that participation involves not only responding to prompts but actively shaping knowledge and interactional roles. Without such attention, moments of resistance or negotiation may be misinterpreted as errors rather than meaningful expressions of agency.

These findings have implications for the design and study of AI-supported learning. Children's displays of knowing extend beyond verbal responses to include embodied and interactional cues, suggesting the need for AI systems that better recognize and respond to these forms of participation. Future work could examine how children negotiate epistemic authority with human versus AI partners, as well as how they themselves perceive these interactions. While based on a small number of cases, this study offers analytic insight into how children seek recognition and position themselves as knowers, informing more responsive and equitable AI designs that support children as active contributors in knowledge building.

Endnotes

Supplementary materials, including a glossary of transcription symbols and conventions, as well as transcript extracts, are available at:
https://docs.google.com/document/d/1Z2jOqtSOuWw0e_EOI85RZBYfAG0X6jy9dpDZSRfV8ec/edit?usp=sharing.

References

Barton, A. C., & Tan, E. (2010). *we be burnin'!*agency, Identity, and science learning. *Journal of the Learning Sciences*, *19*(2), 187–229. https://doi.org/10.1080/10508400903530044

Bautista, J., & Davis, E. A. (2024). Preservice Teachers' support for children's epistemic agency in Elementary Science. *Proceedings of the International Conference of the Learning Sciences*, 19–26. https://doi.org/10.22318/icls2024.669439

Baze, C. L., González-Howard, M., Sampson, V., Crawford, R., & Hamilton, X. (2025). "okay we're doing my idea": How students enact epistemic agency and power in a design-based engineering context. *EducationSciences*, *15*(4), 402. https://doi.org/10.3390/educsci15040402

Carey, S. (2000). Science Education as conceptual change. Journal of Applied Developmental Psychology, 21(1), 13–19. https://doi.org/10.1016/s0193-3973(99)00046-5

Damşa, C. I., Kirschner, P. A., Andriessen, J. E., Erkens, G., & Sins, P. H. (2010). Shared epistemic agency: An empirical study of an emergent construct. *Journal of the Learning Sciences*, *19*(2), 143–186. https://doi.org/10.1080/10508401003708381

Erickson, F. (2006). Definition and analysis of data from videotape: Some research procedures and their rationales. In J. L. Green, G. Camilli, P. B. Elmore, A. Skukauskaiti, & E. Grace (Eds.), *Handbook of complementary methods in Education* research (pp. 177–191). Lawrence Erlbaum Associates.

Goodwin, C. (2013). The co-operative, Transformative Organization of human action and knowledge. *Journal of Pragmatics*, *46*(1), 8–23. https://doi.org/10.1016/j.pragma.2012.09.003

Goodwin, C. (2000). Pointing and the collaborative construction of meaning in aphasia. In *Texas Linguistic* Form (Proceedings of the Seventh Annual Symposium About Language and Society) – Austin SALSA), 43 (pp. 67–76).

Goodwin, C. (2018). *Co-operative action*. Cambridge University Press.

Jefferson, G. (2004). Glossary of transcript symbols with an introduction. *Pragmatics & Beyond New Series*, 13–31. https://doi.org/10.1075/pbns.125.02jef

Jordan, B., & Henderson, A. (1995). Interaction analysis: Foundations and practice. *Journal of the Learning Sciences*, *4*(1), 39–103. https://doi.org/10.1207/s15327809jls0401_2

Jung, Y. J., Whalen, D. P., & Zimmerman, H. T. (2022). Shifts of epistemic agency between children and parents during making at museum-based Makerspace. *International Journal of Science Education, Part B*, *13*(1), 1–17. https://doi.org/10.1080/21548455.2022.2100940

Mitchell, J. C. (1984). Case studies. In R. Allen (Ed.), Ethnographic research: A guide to general conduct (pp. 238–239). London: Academic Press.

Mondada, L. (2018). Multiple temporalities of language and body in interaction: Challenges for transcribing multimodality. *Research on Language and Social Interaction*, *51*(1), 85–106. https://doi.org/10.1080/08351813.2018.1413878

Scardamalia, M. (2002). Collective cognitive responsibility for the advancement of knowledge. *Liberal education in a knowledge society*, *97*, 67-98.

Scardamalia, M., & Bereiter, C. (1991). Higher levels of agency for children in knowledge building: A challenge for the design of New Knowledge Media. *Journal of the Learning Sciences*, *1*(1), 37–68. https://doi.org/10.1207/s15327809jls0101_3

STROUPE, D. (2014). Examining classroom science practice communities: How teachers and students negotiate epistemic agency and learn science-as-practice. *Science Education*, *98*(3), 487–516. https://doi.org/10.1002/sce.21112

Vosniadou, S. (1994). Capturing and modeling the process of Conceptual Change. Learning and Instruction, 4(1), 45–69. https://doi.org/10.1016/0959-4752(94)90018-3

Acknowledgments

This work was supported by the National Science Foundation under Grant No. 2302730. We thank the Ann Arbor District Library for providing space for our study sessions and the families and children who participated.

International Society of the Learning Sciences

A Case Study on the Effect of TAs' Awareness of Pre-class Report Submissions on Scaffolding in Small-group Discussions

Nozomi Fujikawa, Sanii Mizuno, Shintaro Uchiyama, Yusuke Morita
n.fujikawa@aoni.waseda.jp, sunny.mizuno@suou.waseda.jp, uchiyama@aoni.waseda.jp, ymorita@waseda.jp
Waseda University

Abstract: In recent years, Teaching Assistants (TAs) have been expected to play a key role in facilitating small-group discussions. However, prior studies have shown that TAs often struggle to provide appropriate interventions, frequently offering task-related explanations without first diagnosing students' understanding. This study examined whether awareness of students' pre-class report submissions supported TAs' diagnostic scaffolding during small-group discussions. A 100-minute class session was video-recorded, and all TA utterances were transcribed and analyzed based on the framework of the Model of Contingent Teaching (MCT), focusing on the diagnostic and intervention phases. The results suggest that although both TAs were aware of students' pre-class report submissions, only one TA engaged in diagnostic questioning to assess students' understanding before intervening, whereas the other tended to respond with empathic support without diagnosis. This study extends the MCT to TA-supported learning environments by emphasizing the importance of recognizing pre-class reports and highlighting the interpretive processes that mediate between information awareness and adaptive intervention.

Background

The role of the facilitator is important for successful collaborative learning. Teachers ask open-ended questions to promote group progress and support students' metacognitive engagement and collaborative knowledge construction (Hmelo-Silver & Barrows, 2008). During collaborative learning, teachers adapt the level and timing of scaffolding to facilitate social, cognitive, and metacognitive engagement (Kaendler et al., 2015). The Model of Contingent Teaching (MCT), based on Wood et al. (1976), highlights diagnosis as the first step in providing cognitive scaffolding (Van de Pol et al., 2012). Recent studies have further demonstrated that contingent scaffolding effectively promotes students' uptake and understanding when teachers diagnose before intervening (Van de Pol et al., 2019). Moreover, it is important to assess each student's level of understanding within the group before deciding whether to provide collaborative scaffolding or cognitive scaffolding (Calor et al., 2022). Still, teachers face the challenge of insufficient diagnostic strategies to assess students' learning progress during collaborative learning (Bohm et al., 2025).

In recent years, Teaching Assistants (TAs) have been expected to play an important role in supporting students' small-group discussions (Wadams & Schick-Makaroff, 2022). TAs who assist classroom activities alongside teachers can bridge the gap between teachers and students by providing immediate, context-sensitive feedback and recognizing learning difficulties that teachers might overlook (Wilson & Varma-Nelson, 2016). In this regard, it is important for TAs to support students' collaborative learning by providing scaffolding that promotes shared understanding through dialogue. However, most TA–student interactions focused primarily on task-related explanations and problem solving rather than understanding students' thinking, indicating there are challenges with the TAs' scaffolding (D'Angelo & Rajarathinam, 2024). Thus, developing methods that support the diagnostic phase of scaffolding is essential for improving the adaptiveness of intervention.

Understanding students' information beforehand is effective for teachers and TAs in providing appropriate scaffolding during whole-class discussions and collaborative learning. An experienced teacher orchestrated whole-class sense-making discussions by adjusting her in-the-moment pedagogical reasoning according to students' confidence, participation patterns, and classroom relationships (Krist, 2023). Similarly, teachers drew on contextual knowledge—such as knowledge of individual students and of prior lessons—when noticing and interpreting students' strategies in order to support collective reasoning and discussion (Jarry-Shore & Borko, 2025). Fujikawa and Morita (2025) suggested that TAs' awareness of students' pre-class reports can help them decide when and how to intervene appropriately in collaborative learning.

Based on these previous studies, we hypothesized that TAs could use students' submission status of pre-class reports as a diagnostic cue to scaffold effectively during group discussions. Therefore, this study addresses the following research question: Does awareness of pre-class report submissions promote diagnostic scaffolding by TAs? This study extends MCT to TA-supported learning environments by focusing on TAs' use of pre-class information.

 © ISLS

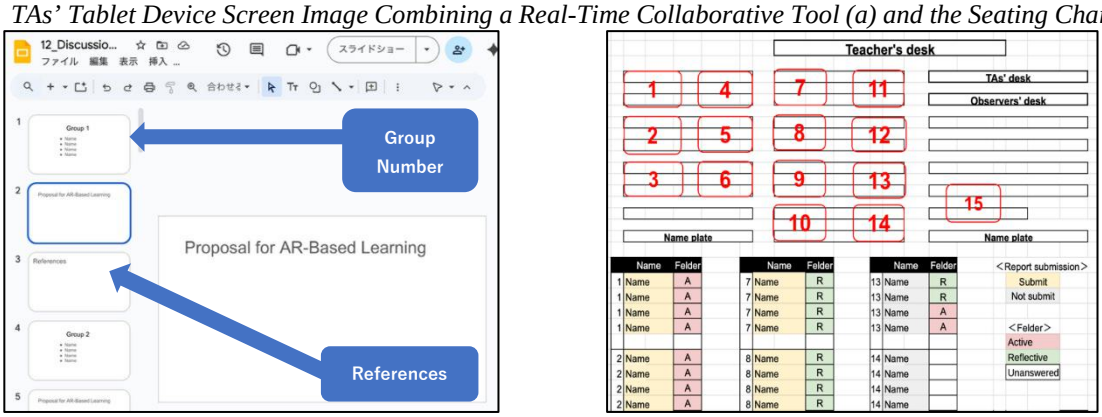
Method

The research field was a class in a Japanese university called “Information Media Education Theory,” conducted from April to July 2025. In this class, students engaged with educational technology research using information media through discussion and presentation. It was a flipped classroom in which students watched preparatory videos, submitted reports, and participated in face-to-face classes. There were 68 second- to fourth-year undergraduate students, one teacher, and two TAs. TA1 was in the second year of a master’s course at a graduate school, and his specialty was educational technology. He had not previously taken or served as a TA for the class. TA2 was a fourth-year undergraduate student, and his specialty was also educational technology. He had experience taking the class but he had no experience acting as a TA for the class. The analysis focused on the 12th session of 14 sessions, which explored new learning methods using AR (Augmented Reality). The discussion topic was to summarize the possibilities and challenges of AR-based learning and propose new learning methods and materials. The discussion time was 60 minutes. In this class, all students were given access to a real-time collaborative editing tool (Google Slides, n.d.) and instructed to summarize the slide numbers and key points of the discussion for each group. Students were instructed to support their ideas with references.

During a one-hour meeting before class, the teacher gave the TAs three instructions. First, the teacher explained the content of the pre-class report, in which students compared VR (Virtual Reality) and AR technologies based on prior research. Second, the first author provided a seating chart indicating groups by report submission status (Submitted groups vs. Non-submitted groups). As shown in Figure 1, the cells of Submitted groups are indicated in yellow and Non-submitted groups are shown in gray. The Index of the Learning Styles (Felder & Supurlin, 2005) is listed to the right of each students’ name, but they were excluded from the analysis this time. Finally, the teacher instructed the TAs to provide support for students while referring to the tool and the seating chart. Both TAs viewed the tool and the seating chart on a tablet device (iPad) during discussions at the same time.

Figure 1

TAs’ Tablet Device Screen Image Combining a Real-Time Collaborative Tool (a) and the Seating Chart (b)



(a) (b)

The data collection and analysis was conducted as follows. First, the discussion utterances of all 14 groups were recorded using a 360-degree camera (PIXPRO SP360 by Kodak), with the exception of one group that did not give consent. Next, the recordings were transcribed verbatim, and all TA utterances were extracted. Utterances referring to report content or submission status were analyzed, and TA interventions were classified. Utterances were categorized into “Diagnostic strategies,” “Checking of diagnosis,” and “Intervention strategies” based on MCT (Van de Pol et al., 2012), with all coding conducted by the first author. Utterances expressing affective support were coded separately as “Empathy.” Finally, we compared the two TAs’ scaffolding based on one submitted and one non-submitted group, analyzing differences in strategies and alignment with the MCT.

Results

During the first 20 minutes of the small-group discussions, both TAs provided interventions that reflected whether they were aware of the students’ pre-class reports. As shown in Table 1, TA 1 demonstrated Diagnostic Strategies, Checking Diagnosis, and Intervention Strategies in the submitted group, and both Diagnostic Strategies and Intervention Strategies in the non-submitted group. On the other hand, TA 2 exhibited intervention utterances in both submitted and non-submitted groups, while no Diagnostic Strategies and Intervention Strategies utterances were observed. In addition, empathic utterances, which were not included in the MCT framework, were observed in TA2’s interactions.

 © ISLS

Table 1

Presence of TAs' Utterances Observed During Scaffolding in Submitted and Non-Submitted Groups

Codes	TA1		TA2	
	Submitted group	Non-submitted group	Submitted group	Non-submitted group
Diagnostic Strategies	+	+	-	-
Checking diagnosis	+	-	-	-
Intervention strategies	+	+	+	+
Empathy	-	-	+	+

Note. + = observed, - = not observed. Empathy was coded separately from the MCT framework.

For Submitted groups, both TAs encouraged students to share what they had found in the literature as a starting point for idea generation. TA1 asked students if they had found any interesting papers during their preliminary reports and advised them to start discussions from there.

Example 1 (TA1– Submitted group)

TA1*:* *Did you find any interesting uses of AR in the prior research?* Like, something you thought was cool? VR is fine too. *Did you each find anything like that?* **[Diagnostic strategies]**

(Students A and C check their computers.)

Student 1: With VR, we can escape the real world, so you could say it's good for practice, we can do anything.

TA1: *With AR, the scenarios are a bit more limited?* **[Checking of diagnosis]**

Student 1: The scenarios are limited…

TA1: That's good. *For example, you each wrote about different things in the report, right? So maybe you can combine those perspectives to come up with ideas.* **[Intervention strategies]**

In contrast, TA2 first showed empathy toward the student before intervening. Although TA2 did not diagnose the situation, he advised the student to proceed with the discussion based on the pre-class report, similar to TA1.

Example 2 (TA2 – Submitted group)

Student 2: With VR, you can actually experience things, so even if it costs more, I think it's worth it. But with AR… I'm not really sure what the benefit is.

Student 3: Yeah, it feels like we're using a new learning method just for the sake of using it.

Student 2: It's just that things become three-dimensional, that's all.

Student 3: Right, so the question is whether we can really understand anything from it.

Student 2: That's what I was thinking while watching the examples.

Student 3: What other problems might there be?

TA2: *Yeah, it's difficult, isn't it?* **[Empathy]**

Student 3: Yeah, it's kind of tricky.

TA2: You know, last time—or was it this *time?—we had you write a short report, right? Try referring to that while writing about the differences from VR. As for the problems, well… it's true that it doesn't necessarily have to be AR. You could write about possible alternatives, or if there are specific reasons why it has to be AR, there are plenty of papers on that. So, look into those gradually.* **[Intervention Strategies]**

For Non-submitted groups, both TAs had advised the groups that had not yet submitted to review the papers first. TA1 discovered that a specific student was researching and sharing papers, diagnosed the situation of other students, and then advised them to research papers.

 © ISLS

Example 3 (TA1 – Non-submitted group)

TA1: *Did you (Student 4) find this?* **[Diagnostic strategies]**

Student 4: It was a paper I found.

TA1: *Did anyone else find anything? Papers or anything?* **[Diagnostic strategies]**

Student 5: No one did the assignment.

TA1: *Try looking into various things right now. AR and VR are quite similar.* **[Intervention Strategies]**

Student 5: They're similar, huh.

TA1: *The professor explained it in a pretty straightforward way.* **[Intervention Strategies]**

Student 5: Should we list the issues that aren't related to VR, or the parts not involved with VR?

TA1: Right. I don't think there's anything not involved; I think everything is connected. The level at which you discuss it will likely emerge as you look at the prior research. *First, try to find what kinds of cases exist for VR and AR.* **[Intervention Strategies]**

Everyone: Yes.

On the other hand, TA2, like the submitted group, first showed empathy and then advised them to review the paper. TA2 provided advice on research methods and key points for discussion and explained why this class has extended discussion time.

Example 4 (TA2 – Non-submitted group)

Student 6: It kind of seems like it overlaps with other groups.

TA2: *It's pretty tough, actually. I struggled with it myself.* **[Empathy]** Anyway, regarding possibilities and issues, they're bound to overlap anyway. I mean, it's not like there are a hundred or two hundred of them. *That part's fine, but for new learning methods, materials, or new learning approaches, try your best to come up with your own unique ideas.* **[Intervention Strategies]**

Student 7: Well, I figured this would come up, right? Like, right about now.

Student 6: We actually had a couple of those.

Student 7: Two? Seriously?

TA2: *That's why, if you come up with a few ideas and feed them to ChatGPT, or if there are papers on the subject, you can research those and tweak them to make something new. Try that approach.* **[Intervention Strategies]** As for feasibility this time, whether it's doable or not, we're leaving that up to the group. So even if it seems a bit unrealistic, it's not the end of the world.

Student 6: If it's interesting, that might be good.

TA2: Interesting is important too. Since it's teaching material, stuff that sparks learner interest probably gets higher approval ratings from students. *That's why we set aside quite a bit of time for discussion.* **[Intervention Strategies]**

Discussion

The results revealed that awareness of students' pre-class reports submissions alone does not necessarily lead to diagnostic scaffolding, as the two TAs used this information differently during group discussions. TA1 engaged in diagnostic processes to infer students' understanding, whereas TA2 tended to provide affective support without diagnosis. This difference may reflect their familiarity with the course: TA1, unfamiliar with the class, inferred students' understanding from dialogue, whereas TA2, a former participant, anticipated difficulties and intervened without diagnosis. Some of TA2's utterances were categorized as empathic responses, suggesting that TAs may sometimes prioritize socio-emotional support over diagnostic scaffolding during group discussions. These findings suggest that awareness of pre-class information alone is insufficient, and that support for translating such awareness into diagnostic reasoning may be necessary. This highlights the importance of supporting TAs' ability to diagnose students' understanding during group discussions. As this study analyzed interactions from a single class session involving two TAs, the findings should be interpreted as exploratory.

 © ISLS

References

Bohm, J., Alrutz, J., & Timm, A. (2025). Adaptive guidance for uncertainty: How teachers use scaffolding in transdisciplinary courses. *Teaching and Teacher Education*, 139, 104461.

Calor, S. M., Dekker, R., van Drie, J. P., & Volman, L. L. (2022). Scaffolding small groups at the group level: Improving the scaffolding behavior of mathematics teachers during mathematical discussions. *Journal of the Learning Sciences*, *31*(3), 369-407

D'Angelo, C. M. & Rajarathinam, R. J. (2024). Speech analysis of teaching assistant interventions in small group collaborative problem solving with undergraduate engineering students. *British Journal of Educational Technology*, 55(4), 1583-1601.

Felder, R. M., & Supurlin, J. (2005). Applications, Reliability and Validity of the Index of Learning Style. *International Journal of Engineering Education*, 21(1), 103–112.

Fujikawa, N., & Morita, Y. (2025). Analysis of Teaching Assistant's Support Process in Small-Group Discussions. In Oshima, J., Chen, B., Vogel, F., & Järvelä, J. (Eds.), *Proceedings of the 18th International Conference on Computer-Supported Collaborative Learning - CSCL202*5 (pp. 563-565). International Society of the Learning Sciences. https://doi.org/10.22318/cscl2025.113215

Google Slides (n.d.). https://workspace.google.com/products/slides/ (last accessed 2025-10-4)

Hmelo-Silver, C. E. & Barrows, H. S. (2008). Facilitating Collaborative Knowledge Building. *Cognition and Instruction*, *26*(1), 48-94

Jarry-Shore, L., & Borko, H. (2025). The role of contextual knowledge in noticing students' strategies in the moment. *Journal of Teacher Education*, 76(1), 89–103.

Kaendler, C., Wiedmann, M., Rummel, N., & Spada, H. (2015). Teacher Competencies for the Implementation of Collaborative Learning in the Classroom: a Framework and Research Review. *Educational Psychology Review*, 27, 505-536.

Krist, C. (2023). Which ideas, when, and why? An experienced teacher's in-the-moment pedagogical reasoning about facilitating student sense-making discussions. *Journal of Research in Science Teaching*, 61(2), 255–288.

Van de Pol, J., Volman, M. & Beishuizen, J. (2012). Promoting teacher scaffolding in small-group work: A contingency perspective. *Teaching and Teacher Education*, 28, 193-205.

Van de Pol, J., Mercer, N., & Volman, M. (2019). Scaffolding student understanding in small-group work: Students' uptake of teacher support in subsequent small-group interaction. *Journal of the Learning Sciences*, *28*(2), 206–239.

Wadams, M. L. & Schick-Makaroff, K. (2022). Teaching assistant development and contributions in online, MOOC and blended synchronous settings: an integrative review. *Journal of Further and Higher Education*, 46(8), 1023-1039.

Wilson, S. B., & Varma-Nelson, P. (2016). Small groups, significant impact: A review of peer-led team learning research with implications for STEM education reform. *Journal of Chemical Education*, 93(10), 1686–1702.

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. J*ournal of Child Psychology and Psychiatry*, 17(2), 89–100.

Acknowledgments

This work was supported by Waseda University Grant for Special Research Projects (2025E-048).

International Society of the Learning Sciences

Reframing Social Annotation as a Text-Embedded Collaboration Analytics Ecology: The Idea Creation Cycle in Social Annotation

Yeonji Jung, Texas A&M University, yjjung@tamu.edu
Sophia Soomin Lee, University of Florida, soomin.lee@ufl.edu

Abstract: While the emergence and evolution of ideas are central to collaborative learning, these processes often remain invisible in digital environments. Social annotation (SA) embeds collaboration within shared texts, forming a distinct ecology where reading, annotation, and discussion occur as a single, integrated workflow within one representational space. This study introduces the Idea Creation Cycle in Social Annotation (ICCSA), a theory-informed and analytics-operationalized framework that models how ideas move and transform through four recursive phases: sensemaking, anchoring, interactive elaboration, and re-anchoring. Using multimodal SA traces, ICCSA captures the temporal trajectories of ideas, examining how cognition, artifacts, and interaction coevolve within one representational space. A proof-of-concept analysis of undergraduate course data reveals distinctive transition patterns and highlights re-anchoring as a key mechanism of knowledge advancement. Together, these findings position ICCSA as a bridge between CSCL theory and actionable learning analytics, advancing the understanding and support of idea development in text-embedded collaboration.

Introduction

In collaborative learning contexts, the emergence of new ideas is critical for knowledge advancement. However, the processes through which ideas emerge and take shape typically remain invisible in conventional classroom practice. Social annotation (SA) offers a unique opportunity to make these processes visible by embedding highlights and comments directly into shared digital texts, allowing learners to externalize their thinking in situ and engage in dialogue around shared anchors in specific textual passages (Kalir & Garcia, 2021; Novak et al., 2012). Accordingly, SA has been widely adopted in higher education, particularly in online and STEM settings, to support both individual comprehension and collaborative knowledge construction (Zhu et al., 2023).

SA is not merely a strategy to increase participation during reading; rather, it intertwines reading, textual anchoring, and peer discussion into a single, text-embedded workflow, which can be viewed as a distinct computer-supported collaborative learning (CSCL) ecology where cognition, artifacts, and interaction coevolve within one shared representational space (Zhu et al., 2020). In this ecology, ideas emerge at the intersection of private comprehension and public inscription. Students select key passages, articulate interpretations or questions in the margins of text, and invite immediate uptake through threaded dialogue. Despite this tightly integrated configuration of sensemaking, anchoring, and interaction, current research often treats these layers separately or reduces analysis to surface indicators focusing on what or how learners contribute (e.g., annotation quantity). While prior studies provide valuable analytical insights into participation and idea creation by proposing conceptual frameworks with operationalized analytic metrics (Zhu & Chen, 2023) and delving into specific reply patterns (Chen & Chen, 2024), those approaches rarely capture how ideas move and transform across these layers of activity in SA.

This study addresses these gaps by introducing the Idea Creation Cycle in Social Annotation (ICCSA), a theory-informed, analytics-operationalized framework that conceptualizes SA as a recursive process of sensemaking, anchoring, interactive elaboration, and re-anchoring/advancement. ICCSA conceptualizes learning as a socio-material cycle in which cognition, artifact, and interaction continuously shape one another, and operationalizes this process through multimodal indicators integrating temporal, semantic, positional, and network dimensions. Specifically, this study (1) operationalizes each ICCSA phase and branching outcome through measurable indicators, (2) constructs an integrated analytic pipeline that identifies phase transitions and idea trajectories, and (3) validates the empirical distinctiveness of these patterns through a proof-of-concept application using undergraduate course data. By consolidating dispersed metrics into a unified, transition-oriented model, ICCSA reframes SA as a first-class context for CSCL theory and provides actionable analytic instruments for modeling and supporting idea development in text-embedded collaboration.

The Idea Creation Cycle in Social Annotation (ICCSA)

The Idea Creation Cycle in Social Annotation (ICCSA) is a *process*-oriented framework that models how ideas emerge, evolve, and advance within text-embedded collaboration. ICCSA positions SA as a distinct socio-material

 © ISLS

ecology in which cognition, artifact, and interaction coevolve within a single, integrated workflow. In this ecology, students interpret shared texts, anchor their insights to specific passages, and build on peers' ideas through threaded dialogue. ICCSA seeks to make these often-invisible movements of ideas empirically visible by linking cognitive, textual, and social layers of collaborative meaning-making to multimodal traces generated by SA learning platforms. Rooted in collaborative discourse and knowledge-building traditions in the CSCL field, ICCSA conceptualizes learning as a process of refining and improving ideas through socially mediated inquiry (Scardamalia & Bereiter, 2014; van Aalst, 2009; Weinberger & Fischer, 2006). ICCSA extends these perspectives by showing how idea movement can be observed in text-anchored contexts where reading, annotation, and discussion are intertwined. Specifically, ICCSA builds on artifact-mediated perspectives in CSCL, foregrounding how material representations shape and sustain collective cognition (Suthers et al., 2010). In SA, the text itself functions as a shared artifact that coordinates comprehension and dialogue, while anchored comments make ideas public and durable, allowing others to elaborate, contest, or recontextualize them. Through this lens, ICCSA fuses cognitive and social processes that are often studied separately into a recursive cycle of externalization and transformation.

Additionally, temporal and sequential perspectives on CSCL (Popov et al., 2017; Rosé et al., 2020) further inform ICCSA's phase-transition view of idea development. These approaches emphasize that learning unfolds as interdependent actions whose order and timing shape meaning-making opportunities. Rather than treating annotations as isolated events or aggregated participation metrics, ICCSA models ordered sequences of actions that trace the life cycle of an idea, from private cognitive preparation (sensemaking), to public textual externalization, to social elaboration and subsequent recontextualization. Through this view, ICCSA advances recent analytical efforts that model collaborative discourse through networked perspectives, such as the Attention-Interaction-Creation (AIC) (Zhu & Chen, 2023) and Social Semantic Network Analysis (Oshima et al., 2019). While these models reveal structural and relational patterns of SA (Zhu & Chen, 2023) or examine temporal dynamics of idea networks in broader CSCL contexts (Ohsaki & Oshima, 2021), ICCSA shifts the analytic focus from structural connectivity to temporal transitions across phases of idea work, situating these dynamics in the distinctive ecology of SA where the text itself organizes participation. By integrating behavioral (e.g., reading durations), textual (e.g., semantic continuity), and social (e.g., thread branching) indicators, ICCSA establishes a conceptual-behavioral-analytic bridge between CSCL theory and computationally observable traces. This enables empirical modeling of how ideas are initiated, elaborated, and re-anchored over time within the ecology of text-embedded collaboration.

ICCSA delineates four recursive phases of idea creation (sensemaking, anchoring, interactive elaboration, and re-anchoring), each representing a distinct coupling of cognition, artifact, and interaction within SA ecology. Learners first make sense of text (sensemaking), externalize ideas as epistemic artifacts (anchoring), refine them through social negotiation (interactive elaboration), and recontextualize them into new spaces (re-anchoring). Each phase generates distinct but interlinked behavioral and semantic traces, enabling analytics to capture transitions (e.g., anchoring → elaboration → re-anchoring versus anchoring → dormancy). Through this process view, ICCSA connects micro-level annotation behaviors to macro-level trajectories of idea development, providing a framework for both theoretical explanation and computational modeling of collaborative sensemaking in SA.

Phase 1: Sensemaking. Sensemaking precedes visible contribution and captures how learners allocate attention, interpret text, and prepare to externalize ideas. The text functions as a cognitive scaffold, prompting the identification of ambiguities or connections that later become public annotations. Drawing on theories of joint attention and preparatory comprehension (Chi, 2009; Wise et al., 2021), ICCSA operationalizes this "invisible" cognitive work through behavioral indicators such as reading duration, scrolling coverage, and highlight density before annotation (see Table 1). These traces reveal learners' investment in comprehension and the establishment of shared attentional context, which is an essential foundation for subsequent anchoring.

Phase 2: Anchoring. Anchoring marks the transition from private understanding to public articulation. In this phase, learners externalize emerging ideas by attaching them to specific textual passages through highlights and annotations, creating visible, text-bound epistemic artifacts that can be taken up by others (Scardamalia & Bereiter, 2006; Suthers, 2006). The highlighted passage and its associated annotation together function as a boundary object linking individual interpretation with collective discourse. Conceptually, anchoring is defined by how ideas are first made visible and situated within the shared representational space. Accordingly, analytic indicators focus on two dimensions: (1) the specificity of text linkage (captured through highlight span length and anchor precision, reflecting how tightly an idea is localized to a textual segment) and (2) the form of initial externalization (captured through annotation length, syntactic completeness, and epistemic stance such as questioning, which characterize how learners position their contributions at the moment of anchoring). Together, these indicators capture how ideas enter the collaborative space and enable subsequent elaboration.

Phase 3: Interactive Elaboration. Interactive elaboration forms the social core of ICCSA, where anchored ideas are tested, refined, or restructured through dialogue. In this phase, learners engage with others' contributions by building on, questioning, or contesting prior annotations, reflecting transactive discourse and dialogic engagement (Weinberger & Fischer, 2006; Stahl et al., 2014). Compared to anchoring, which focuses on initial inscription, interactive elaboration captures how ideas evolve through interaction over time. Analytic indicators focus on two dimensions: (1) the structure of interaction and (2) the semantic progression of ideas. Structural indicators include thread depth, branching factor, and reply latency, reflecting the extent, responsiveness, and sustainability of dialogue. Semantic indicators capture how ideas develop within threads, including semantic cohesion, stance shifts, and the introduction of new perspectives. Through these threaded interactions, ICCSA identifies distinct trajectories of idea development, including divergence (where peers introduce novel perspectives and expand conceptual space), convergence (where dialogue consolidates shared understanding), discarding (where ideas are refuted or abandoned), and dormancy (where anchors receive little or no uptake), revealing how collaborative engagement shapes ideas beyond their initial articulation.

Phase 4: Re-Anchoring/Advancement. Re-anchoring represents the recursive culmination of the cycle, where learners revisit and extend prior ideas across time and contexts. In this phase, ideas are not only elaborated within a single thread but recontextualized through new annotations that connect back to earlier contributions, either within the same text or across texts. This process operationalizes idea advancement in knowledge-building terms, reflecting the continued refinement and reuse of ideas over time (Scardamalia & Bereiter, 2014; Stahl, 2017). Analytic indicators focus on the continuity and integration of ideas across contexts. Semantic continuity captures the degree to which new annotations relate to prior ones, indicating reuse or transformation of earlier ideas. Cross-text linkage and re-anchoring latency reflect how ideas travel over time. Network-based indicators such as bridging centrality capture how learners connect otherwise separate discussion clusters. Together, these measures reveal how ideas accumulate, integrate, and advance, highlighting the role of learners in sustaining knowledge development within text-embedded collaboration.

Table 1

Operationalization of ICCSA Phases and Analytic Metrics

Phase	Metric	Definition / Computation	Data Source	Theoretical Construct
Phase 1: Sensemaking	Pre-annotation Duration	Time between document access and first annotation	Reading + first annotation timestamp	Cognitive investment and preparatory comprehension before visible contribution
	Inter-annotation Interval	Mean time gap between consecutive annotations authored by the same learner	Annotation timestamps	Reflective rereading and cognitive spacing
Phase 2: Anchoring	Highlight Span Length	Number of words or characters in highlighted passage	Highlight spans	Specificity of text linkage
	Anchor Precision	Ratio of highlight length to surrounding text unit	Highlight spans + document structure	Boundedness of idea-text linkage
	Annotation Length	Token count of annotation text	Annotation content	Degree of initial externalization
	Annotation Completeness	Presence of full syntactic structure or interpretive markers	Annotation text	Coherence of epistemic artifact
	Annotation Type / Stance	Classification (e.g., question, interpretation, critique)	Annotation content	Initial epistemic orientation
Phase 3: Interactive Elaboration	Thread Depth	Maximum reply nesting per thread	Reply structure	Sustained dialogic engagement
	Branching Factor	Number of replies per node	Reply structure	Expansion of interaction
	Reply Latency	Time between anchor and first reply	Timestamps	Responsiveness and uptake
	Semantic Cohesion	Similarity among annotations within a thread	Text embeddings	Conceptual convergence
	Stance Shift	Change in discourse stance across replies	NLP classification	Epistemic transformation
Phase 4: Re-anchoring/ Advancement	Semantic Continuity	Similarity between new and prior annotations	Embeddings + temporal order	Idea reuse and development
	Cross-text Linkage	Connections across different readings or documents	Annotation + text IDs	Transfer and recontextualization
	Re-anchoring Latency	Time between original and linked annotations	Timestamps	Temporal extension of ideas
	Network Bridging	Betweenness centrality in interaction-semantic network	Multilayer network	Integration across discussion clusters

 © ISLS

Proof-of-concept: Applying ICCSA in an undergraduate online course

To demonstrate the empirical applicability of the ICCSA framework, this study conducted a proof-of-concept analysis using data from an undergraduate-level online asynchronous course in statistics that integrated SA as a core learning activity. The course was designed around weekly readings, during which students collaboratively read, annotated, and discussed shared scholarly articles using a digital annotation platform named Perusall. The dataset comprised de-identified Perusall exports containing reading times, highlight spans and positions, annotation texts, timestamps, and reply structures, enabling reconstruction of students' multimodal engagement across cognitive, textual, and social layers of discourse. For analytic consistency and comparability, we selected the same Week 5 reading used in both the 2022 and 2023 course iterations. This midpoint reading provided a balanced volume of activity, stable instructional context, and structured pagination that allowed precise data alignment. The proof-of-concept focused on three analytic objectives: (1) operationalizing each ICCSA phase through measurable multimodal indicators; (2) constructing an integrated analytic pipeline to detect temporal transitions among phases; and (3) examining cross-year differences to validate the empirical distinctiveness and robustness of the framework. Each ICCSA phase was instantiated using theoretically grounded metrics derived from Perusall logs: temporal (e.g., reading and inter-annotation intervals), textual (e.g., annotation length, semantic continuity), interactional (e.g., thread depth, reply latency, cohesion), and network-based (e.g., bridging centrality, cross-thread linkage). These multimodal traces were processed through a unified analytic workflow to generate comparable indicators across cohorts.

Across both cohorts, ICCSA captured the full life cycle of idea activity from preparatory reading to anchoring, elaboration, and re-anchoring. Descriptive contrasts between the 2022 (n = 87) and 2023 (n = 85) classes highlighted both continuity and variation in idea dynamics. In the *sensemaking* phase, 2022 students exhibited longer pre-annotation reading times (M = 14.77 min) and broader coverage, suggesting deeper preparatory engagement than the 2023 cohort (M = 2.38 min). The *anchoring* phase similarly showed richer externalization in 2022, with longer annotations (M = 185 words vs. 125) and greater interpretive distance between annotations and source text (0.70 vs. 0.61). During *interactive elaboration*, the 2022 cohort demonstrated more sustained discussion (thread depth = 1.48 vs. 1.06) and faster uptake (reply latency = 1.47 days vs. 2.55), aligning with higher dialogic engagement. Finally, *re-anchoring* patterns suggested stronger integrative movement in 2022, reflected in slightly higher semantic continuity (0.78 vs. 0.76) and greater network bridging (0.002 vs. 0.0004). Together, these indicators illustrate how ICCSA can identify distinct temporal and structural dynamics of idea development even within the same instructional design.

Discussion and implications

This study introduced ICCSA as a theory-informed and analytics-operationalized framework for examining how ideas emerge, evolve, and advance within text-embedded collaboration. The proof-of-concept analysis demonstrated ICCSA's potential to capture the temporal and structural dynamics of idea development, revealing how cognitive, textual, and social layers interact throughout the collaborative process. By linking these layers to multimodal traces, ICCSA helps move learning analytics toward modeling the processual movement of ideas across initial sensemaking, anchoring, elaboration, and re-anchoring. Theoretically, ICCSA contributes to CSCL theory by framing SA as a socio-material ecology where cognition, artifacts, and interaction coevolve within a single representational space. While prior frameworks of knowledge building and artifact-mediated cognition examined these dimensions separately (Zhu & Chen, 2023; Chen & Chen, 2024) or broader CSCL contexts (Ohsaki & Oshima, 2021), ICCSA unites them into a recursive cycle of externalization and transformation, elucidating how ideas transition from private comprehension to public inscription, from individual interpretation to collective negotiation, and from situated discussion to cumulative knowledge advancement. The inclusion of re-anchoring extends knowledge-building theory by identifying how learners revisit and repurpose earlier insights to sustain inquiry over time. Methodologically, ICCSA introduces a phase-transition approach linking CSCL theory with measurable multimodal analytic indicators. SA data, traditionally analyzed via structural metrics, are reconceptualized as traces of idea trajectories. Behavioral, textual, and social dimensions jointly represent cognitive investment, epistemic stance, and dialogic exchange (e.g., reading time and coverage reflect sensemaking; annotation length, stance, and text-annotation alignment capture anchoring). The proof-of-concept findings show that ICCSA can reveal subtle temporal and structural variations in idea movement that remain invisible in static engagement metrics, offering a shared analytic vocabulary that bridges conceptual and computational models. Together, by linking CSCL constructs to empirically observable patterns in learners' annotation activity, ICCSA provides both a theoretical account of idea dynamics and a basis for analytics-based modeling, supporting the growth and continuity of ideas within the evolving ecology of text-embedded collaborative learning.

 © ISLS

References

Bereiter, C., & Scardamalia, M. (2014). *Knowledge building and knowledge creation: One concept, two hills to climb. In Knowledge creation in education* (pp. 35-52). Springer.

Brown, A. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *The Journal of Learning Sciences, 2*(2), 141–178.

Chen, B., & Chen, Z. (2024). Why do students reply? Uncovering the socio-semantic entanglement in web annotation activities. *Education and Information Technologies, 29*(8), 9747-9771.

Chi, M. T. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science, 1*(1), 73-105.

Kalir, R., & Garcia, A. (2021). *Annotation*. MIT Press.

Lave, J. (1987). *Cognition in practice*. Cambridge University Press.

Novak, E., Razzouk, R., & Johnson, T. E. (2012). The educational use of social annotation tools in higher education: A literature review. *The Internet and Higher Education, 15*(1), 39-49.

Ohsaki, A., & Oshima, J. (2021). Knowledge Creation Analytics for Jigsaw Instruction: Temporal Socio-Semantic Network Analysis. In de Vries, E., Hod, Y., & Ahn, J. (Eds.), *Proceedings of the 15th International Conference of the Learning Sciences - ICLS 2021*. (pp. 537-540). Bochum, Germany: International Society of the Learning Sciences.

Oshima, J., Oshima, R., & Matsuzawa, Y. (2012). Knowledge building discourse explorer: A social network analysis application for knowledge building discourse. *Educational Technology Research and Development, 60*(5), 903-921.

Oshima, J., Oshima, R., Ohsaki, A., & Splichal, J. (2019). Collective Knowledge Advancement through Shared Epistemic Agency: Socio-Semantic Network Analyses. In Lund, K., Niccolai, G. P., Lavoué, E., Hmelo-Silver, C., Gweon, G., & Baker, M. (Eds.), *Proceedings of the 13th International Conference on Computer-Supported Collaborative Learning - CSCL 2019*, Volume 1 (pp. 57-64). Lyon, France: International Society of the Learning Sciences.

Popov, V., van Leeuwen, A., & Buis, S. C. (2017). Are you with me or not? Temporal synchronicity and transactivity during CSCL. *Journal of Computer Assisted Learning, 33*(5), 424-442.

Rosé, C. P., & Järvelä, S. (2020). Building community together: Towards equitable CSCL practices and processes. *International Journal of Computer-Supported Collaborative Learning, 15*(3), 249-255.

Scardamalia, M., & Bereiter, C. (2006). *Knowledge building*: *Theory, pedagogy, and technology*. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 97-118). Cambridge University Press.

Stahl, G. (2017). Group practices: A new way of viewing CSCL. *International Journal of Computer-Supported Collaborative Learning, 12*(1), 113-126.

Stahl, G., Cress, U., Ludvigsen, S., & Law, N. (2014). Dialogic foundations of CSCL. *International Journal of Computer-Supported Collaborative Learning, 9*(2), 117-125.

Suthers, D. D. (2006). Technology affordances for intersubjective meaning making: A research agenda for CSCL. *International Journal of Computer-supported collaborative learning, 1*(3), 315-337.

Suthers, D., & Medina, R. (2010). Tracing interaction in distributed collaborative learning. In *Analyzing interactions in CSCL: Methods, approaches and issues* (pp. 341-366). Springer.

van Aalst, J. (2009). Distinguishing knowledge-sharing, knowledge-construction, and knowledge-creation discourses. *International Journal of Computer-Supported Collaborative Learning, 4*(3), 259-287.

Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & Education, 46*(1), 71-95.

Wise, A. F., Knight, S., & Shum, S. B. (2021). Collaborative learning analytics. In *International handbook of computer-supported collaborative learning* (pp. 425-443). Springer.

Zhu, X., & Chen, B. (2023). Understanding idea creation in collaborative discourse through networks: The joint attention–interaction–creation (AIC) framework. In Damșa, C., Borge, M., Koh, E., & Worsley, M. (Eds.), *Proceedings of the 16th International Conference on Computer-Supported Collaborative Learning - CSCL 2023* (pp. 155-162). International Society of the Learning Sciences.

Zhu, X., Chen, B., Avadhanam, R. M., Shui, H., & Zhang, R. Z. (2020). Reading and connecting: Using social annotation in online classes. *Information and Learning Sciences, 121*(5/6), 261-271.

Zhu, X., Shui, H., & Chen, B. (2023). Beyond reading together: Facilitating knowledge construction through participation roles and social annotation in college classrooms. *The Internet and Higher Education, 59*, 100919.

International Society of the Learning Sciences

Embodied and Non-Embodied Technologies in Mathematics Learning: A Comparative Study

Injila Rasul, Ivon Arroyo
irasul@umass.edu, ivon@cs.umass.edu
University of Massachusetts Amherst

Abstract: WearableLearning is a multiplayer, game-based digital learning environment in which students use mobile devices and collaborate while physically engaging with mathematical concepts such as geometry. This study examines how students learn mathematics in embodied games enabled by WearableLearning compared with a traditional seated digital learning environment. We report results from a randomized controlled trial conducted in a U.S. afterschool program, showing that students with lower prior knowledge benefited the most from the technology-supported embodied learning condition.

Introduction and background

Much research has highlighted embodied learning as a powerful pedagogical approach for introducing and developing mathematical thinking skills (Nathan, 2021; Flood, Shvarts & Abrahamson, 2020). At the heart of embodied learning, the notion of embodied cognition asserts that knowledge is built from a continuum between the mind, body, and the physical world around us, suggesting that incorporating sensorimotor processes in lessons will strengthen learning through the learners' use of their bodies, reflection, and social interaction (Ma, 2017; Georgiou & Ioannou, 2019). Embodied cognition posits that learning involves new ways of moving and interacting with the world around us and the artifacts within our environment, involving figuring out how to move differently, following from a new way of perceiving the environment, using artifacts that can improve our "perception for action" and extend our perception–action loops (Abrahamson et al., 2024). By using these artifacts, we adopt their inherent mechanisms of disciplinary practices, consisting of gestures, discourse, and actions. Here, knowing, as woven in with doing, manifests itself as physical actions that extend the notion that these actions are thinking made visible (Nemirovsky et al., 2012).

Learning Technologies for mathematics widely used in schools, such as Intelligent Tutoring Systems (ITS), have traditionally been screen-based, with students sitting down in front of a computer in isolation from each other, and do not encourage learners to engage in embodied learning practices and typically uses chat windows for collaboration (Rasul et al., 2023a; Jeong & Hmelo-Silver, 2016). Given greater evidence supporting the importance of incorporating embodiment into pedagogy, there is also a need for new learning technologies and for developing ways to support learners through embodiment (Abrahamson et al., 2020; Abrahamson et al., 2024; Hostetter & Alibali, 2008).

With this in mind, the Advanced Learning Technologies Lab, with support from the NSF, has developed a novel learning technology called **WearableLearning (WL)** that promotes K–12 mathematics thinking and learning through physically active (both tangible and embodied) multiplayer math games on mobile devices. Prior research demonstrates improved performance on math tests after engaging with embodied physically active math games using the WL platform (Arroyo et al., 2020; Arroyo et al., 2021; Rasul et al., 2023a; Crabtree et al., 2024). However, there is a lack of controlled studies where the main difference between conditions is the embodiment within activities. This paper compares two learning conditions to answer the question: *To what extent do students learn mathematics differently through tangible and embodied activities facilitated by WearableLearning compared to traditional learning technology that teaches the same mathematical concepts?*

The Common Core State Standards for Mathematics (CCSSM) and Mathematical Practice (SMP) emphasize conceptual understanding, real-world application, and collaborative problem solving. They highlight how math knowledge develops through social interaction, shared reasoning, and engagement with meaningful models. Sensory-motor actions, gestures, and interactions play a key role in learning, as cognition is distributed across the mind, body, and social context (Goldin-Meadow et al., 2009; Congdon et al., 2017). Students also learn by manipulating tangible objects and engaging face-to-face with peers and teachers (Abrahamson, 2024).

Yet, most learning technologies are static (Cook et al., 2016), and do not employ embodied learning, which moves the interactions, at least partly, "off the screen", engaging in hands-on socially rich activities Embodied design principles are a lens for examining how digital resources are created and brought to life in mathematics learning tasks. For instance, Klemmer et al. (2006) reiterate that themes of embodied interaction design are central to understanding how the physical and social contexts of human action influence human-computer interaction. These essential guiding themes are: a. Thinking Through Doing: embodied interaction

 © ISLS

highlights physical engagement in creative and problem-solving processes, b. Performance: incorporating an artifact such that you act through it rather than on it, c. Visibility: Making users' actions and intentions visible to others fosters better collaboration and coordination, d. Risk: Physical action carries a certain risk in terms of possible harm, which follows from making a choice that may not be simple to undo, e. Thick Practice: action is situated in specific social, physical, and cultural contexts. Yet, Melcer and Isbister (2016) realized that embodied games and simulations have utilized a large breadth of design approaches that often have resulted in the creation of seemingly unrelated systems. Their design framework aggregates different design approaches: 1) physical interaction, classified as: embodied, enacted, manipulated, surrogate, or augmented; 2) social interactions that the activity requires (e.g., collaborative); and 3) the role of the environment (e.g., physical/virtual/mixed world).

WearableLearning

WearableLearning (WL) is a web-based platform accessed via mobile devices (e.g., smartphones or tablets) that enables the incorporation of embodied gameplay into classrooms, curricula, and standards for K-12 students (Arroyo et al., 2017; Micciolo et al., 2018). Through WL (wearablelearning.org), students engage in active, collaborative, gamified math activities, exploring physical environments, manipulating tools, interacting with peers and teachers, and internalizing learning through embodiment. Lightweight, portable devices act as "virtual assistants," guiding students and providing feedback as they navigate physical spaces and objects. Targeting the NCC, math game playing and creation with WL has been implemented with roughly 500 K–12 students, in classrooms, summer programs, and afterschool programs, as well as with 25 math teachers, who also have created a variety of games for students (Rasul et al., 2023a; Rasul et al., 2023b; Smith et al., 2020). Test scores showed improvement in math problem-solving across different areas of math and grade levels, despite short 30–40 minute exposures (Rasul et al., 2023a; Arroyo et al., 2021; Crabtree et al., 2024). All games led to learning gains, and students showed high engagement while playing, leading to implications for extending embodied learning through AI support (Rasul et al., 2023b; Arroyo et al., 2024).

WearableLearning: Tangrams race

The Tangrams Race game consists of a relay race that embraces the tangible "manipulated" form of embodied interaction (Melcer & Isbister, 2016), which regards embodied and tangible interactions using mathematically meaningful manipulatives as students learn mathematics, using hand-based manipulations, where physical objects highlight mathematical properties. A tangram is a traditional Chinese puzzle consisting of seven flat geometric pieces that can be arranged without overlap to form various shapes, recreating specific designs. The Tangrams Race requires each team to run from a start line to an end line with a basket full of geometric shapes, a subset of which are the Tangrams shapes. The mobile device (worn on the arm) provides each student various mathematical descriptions of geometric objects that they are asked to retrieve, so they can collaborate to create a Tangram arrangement at the starting line. When students find the correct shape, at the end line, they tell the phone they have found it by inputting an identifying color code (e.g. "red, green, black"), which is on each shape, into the WearableLearning (WL) app. WL provides them with immediate feedback (correct, incorrect, hints when requested), until the students collect the correct piece. Figure 1 shows how the teacher, standing at the end line, gestures with hands when students might get stuck to help students find the correct shape. Students run 3 times in each level to retrieve a geometric shape. Once all the pieces have been collected, the App tells them to work together to rearrange the pieces to replicate the desired pattern (e.g. a boat, a house, a cat) . This game has two levels, with the second being more challenging.

Figure 1

The Experimental Condition (left) vs. the Control Condition (right)



From a cognitive perspective, the Tangrams Race involves a variety of cognitive processes. At the starting line, students are provided a mathematical description of a shape, which encourages visual imagery. After running and arriving at the end line, students search through a basket full of shapes, picking up pieces to consider, physically rotating them, sensing their salient attributes such as vertices and edges, and going through a process of discerning and discarding. If students enter the code for an incorrect shape, this generates cognitive conflict that can be solved in several ways: a) students might give a try to another candidate piece; b) students can ask for

 © ISLS

"hints" by pushing a special "hint" button; c) students may receive support from a teacher stationed at the end line, who interacts with them by asking what piece they are searching for, and who uses hand gestures to illustrate relevant properties, guiding them toward the correct geometric shape without directly revealing the answer. When all nine shapes are retrieved (three per player, and two extra pieces are simply extra distractors), the team works together to compose the 7-piece Tangram shown on their device. Note that *running* to the end line adds excitement and student engagement to the game but is not considered part of the embodied cognitive activities.

Methodology

Procedure and participants

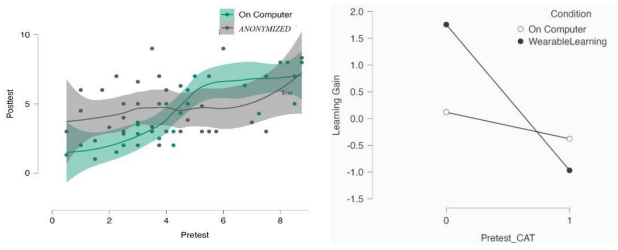
We ran a controlled study in which participants were randomly assigned to the WL (n = 37) and On-Computer (n = 26) groups (see Figure 1). The study participants were 63 students in Grades 4 and 5 enrolled in an afterschool program in the United States. These students belonged to low-socioeconomic backgrounds and historically minority communities, with regional demographics being 48% Hispanic, 29% White, and 18% Black. Students participated in the experience after their school day, typically starting around 4 pm. The Experimental Condition (left of Fig. 1) played the Tangrams Race in WL. Students used phones on their wrists to play, and retrieved geometric pieces to build the Tangram, which were color-coded. Game facilitators naturally gestured at the end line, showing parallel lines, right angles, perpendicular lines, etc. The Control Condition (on the right in Fig. 1) used a digital learning technology that delivered the same math problems and hints as the Tangrams Race. Towards the end of the session, they engaged with a Virtual Tangram, where students rotated and placed geometric pieces with the mouse. Exposure time was about 45 minutes. All participants took a pre- and post-test. The students were tested on 10 content-related questions from a 4th-grade geometry standardized exam in the United States (MCAS) on day one of their participation in the after-school program, following which they played the game and were tested again afterward. The questions chosen reflected the concepts that the game required the students to solve, thereby indicating the conceptual starting point of each student and their subsequent improvement or lack thereof. Using two tests, A and B, the pre-post test distribution was counterbalanced: half of the students received test A for their pretest, and the other half received test B; then, group A's pretest was group B's posttest, and vice versa. Counterbalancing ensured that if one test was unintentionally more difficult than the other, the difference would cancel out, preventing bias in our measure of learning. Data collection for this research began as a graduate thesis (Agbaji, 2019) and continued over multiple years, adding data incrementally to reach the required participants and achieve statistical power.

Results

There was a mean difference of +.39 and a standard deviation of 2.12 from pre-test to post-test across both conditions, indicating that students improved overall from pre-test to post-test, with the experimental group improving more. Yet, there was not a significant main effect for condition (F = 1.17, p = 0.28) even though the effect size was $d = 0.38$ in favor of the experimental condition (see Figure 3, Table 1). Overall, *TangramRace* students had a mean learning gain of 0.75 points (Pretest mean = 3.8, Post-test mean = 4.6), while On-Computer students showed a slight learning loss, averaging -0.16 points (Pretest mean = 4.73, Post-test mean = 4.57).

Figure 2

Left: Score (Pretest to Posttest) Right: Scores by Pretest Score and Condition, and their Learning Gai



Notably, when considering student achievement level (pretest scores indicating math ability prior to the intervention) as a factor, we found a *significant interaction effect* for Condition x Pretest Math Ability (F = 5.32, $p < 0.025$), suggesting that it was lower math ability students who benefited significantly more from the *embodied TangramsRace* condition (see Figure 2). Bonferroni contrasts confirmed that the significant interaction effect was driven by low-math-ability students performing better in the *TangramsRace* than their peers in the control

 © ISLS

condition (see Figure 2), indicating that the embodied TangramsRace condition was more beneficial for students with lower math ability. High-achieving students did not improve in either condition.

Table 1
ANOVA Test Results

	Sum of Squares	df	Mean Square	F	p
Playing Condition	3.9	1	3.9	1.177	0.282
Pretest Score	37.6	1	37.6	11.096	0.001
Condition x Pretest Math Ability	18.0	1	18.0	5.319	0.025
Residuals	200.0	59	3.3		

Discussion

We believe that, in general, the means are deflated due to student fatigue at the end of the school day (posttests were administered around 5 pm). Despite this, students performed better on the post-test than on the pre-test when exposed to *TangramsRace*, yet there was not a significant difference (main effect) between the experimental and control conditions. However, we found a significant interaction effect that indicated students of lower prior knowledge learned most with the embodied WL *TangramsRace* condition than the control condition ($p < 0.025$). Students in the control group were confined to their seats, limited by the stationary setup. Although they solved more problems overall and interacted frequently with the virtual Tangrams manipulative, their learning gains were smaller compared to those in the embodied condition. One explanation is that lower-achieving students have a greater need to build conceptual understanding of geometric shapes, and that the embodied condition supported this through its tangible, sensory-motor nature, which made the relevant mathematical properties more salient. This finding is important as it underscores the potential of embodied technologies to support struggling students.

Tech-supported mobilization supports teacher/peer interactions and provides instant feedback. Rich interactions beyond the digital interface expand opportunities for mathematical thinking through embodied means. This aligns with Abrahamson et al.'s (2024) ecological dynamics framework, which emphasizes grounding new concepts in the perceived affordances of physical mathematical objects and movement.

A second key difference lies in the type and extent of support that students receive from traditional learning technologies compared to WearableLearning. As stated previously, the affordances of traditional learning technologies are well known, and include immediate feedback, hints/supports, assessment data for teachers, and personal pacing through the material. While both traditional learning technologies and WL incorporate these features, closer examination reveals a distinct difference in how students experience support. Given that WL expands the range of interactions students engage in, support extends beyond the digital learning environment and teacher scaffolding. Students benefit from multi-pronged assistance: teacher support, peer support, and some collaborative interactions within their teams. Furthermore, students learning math with WL also have the advantages of tangible support—externally through hands-on interaction with physical materials, and internally through gestures and bodily engagement that help them make sense of mathematical concepts. We claim that these differences between traditional learning technologies and WL account for the differences in quantitative gains in learning, but also in qualitative observations on how students are seen learning math. Yet, further investigation is required to determine *how* these differences in the received support impact student learning.

Conclusion

Our goal was to explore how much students learn with WL as compared to traditional static learning technology. The findings from implementing WL into an afterschool program for math, in two groups–a control condition in a traditional on-computer setting and a physically/socially active condition, demonstrate that embodied learning, using action and gesture, can yield higher gains, particularly for those with lower prior knowledge.

A primary limitation of this study was that it took place in an after-school program at the end of the day, when students were either tired or being picked up by their parents during the lesson, which added noise to the data. Also, the participants were young students (3rd–5th graders). Yet, these results show that: a) it is possible to create equivalent digital math game-like conditions (digital embodied or digital sit-down); b) math ability might be a mediating factor influencing the effectiveness of embodied math tasks. New questions arise for future research: a) Does this same effect extend to other mathematics topics? b) Since the embodiment was related to hand-based manipulations of mathematical objects, does this effect extend to full-scale body embodiment (e.g., walking number lines or being fully immersed in coordinate planes, embodying points on a plane, or enacting transformations)? d) How can we assess mathematics performance in a way that is not 'sitting down' to take a test, but also to assess in an embodied manner? These are questions we intend to explore in future research.

 © ISLS

References

Abrahamson, D., Nathan, M. J., Williams-Pierce, C., Walkington, C., Ottmar, E., et al. (2020). The future of embodied design for mathematics teaching and learning. *Frontiers in Education, 5*, 147..

Abrahamson, D., Tancredi, S., Chen, R. S., Flood, V. J., & Dutton, E. (2024). Embodied design of digital resources for mathematics education: Theory, methodology, and framework of a pedagogical research program. In *Handbook of digital resources in mathematics* education (pp. 217-250). Cham: Springer Int. Publishing.

Agbaji, E. (2019). *Improving math learning with embodied game-based mobile and wearable technologies.* (Unpublished master's thesis). Worcester Polytechnic Institute.

Arroyo, I., Castro, F., Smith, H., Harrison, A., Ottmar, E. (2021). *Augmenting Embodied Mathematics Classrooms with Mobile Tutors.* AERA 2021. https://par.nsf.gov/biblio/10283183.

Arroyo, I., Micciollo, M., Casano, J., Ottmar, E., Hulse, T., & Rodrigo, M. M. (2017). Wearable learning: multiplayer embodied games for math. *Proceedings of 4th ACM SIGCHI Annual Symposium on Computer-Human Interaction in Play*, CHI PLAY 2017. ACM/IEEE.

Arroyo, I., Rasul, I., Crabtree, D., Castro, F., Poh, A., Gattupalli, S., ... & Micciolo, M. (2024, July). Aligning AIED systems to embodied cognition and learning theories. In International Conference on Artificial Intelligence in Education (pp. 3-17). Cham: Springer Nature Switzerland.

Arroyo, I., Woolf, B. P., Burelson, W., Muldner, K., Rai, D., & Tai, M. (2014). A multimedia adaptive tutoring system for mathematics that addresses cognition, metacognition and affect. International Journal of Artificial Intelligence in Education, 24, 387-426.

Congdon, E. L., Novack, M. A., Brooks, N., Hemani-Lopez, N., O'Keefe, L., & Goldin-Meadow, S. (2017). Better together: Simultaneous presentation of speech and gesture in math instruction supports generalization and retention. Learning and instruction, 50, 65-74.

Cook, S.W., Friedman, H.S., Duggan, K.A., Cui, J., & Popescu, V. (2016). Hand gesture and mathematics learning: Lessons from an avatar. Cognitive Science, 1-18.

Crabtree, D., Rasul, I., & Arroyo, I. (2024). Implementing WearableLearning in the math classroom: an exploration into the affordances of embodied learning through game-play compared to traditional learning technologies. In Proceedings of ICLS 2024, pp. 530-537. Int. Society of the Learning Sciences.

Georgiou, Y., & Ioannou, A. (2019). Embodied learning in a digital world: A systematic review of empirical research in K-12 education. Learning in a digital world: Perspective on interactive technologies for formal and informal education, 155-177.

Goldin-Meadow, S., Cook, S. W., & Mitchell, Z. A. (2009). Gesturing gives children new ideas about math. Psychological science, 20(3), 267-272.

Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. Psychonomic bulletin & review, 15(3), 495-514.

Jeong, H., & Hmelo-Silver, C. E. (2016). Seven affordances of computer-supported collaborative learning: How to support collaborative learning? How can technologies help?. Educ. Psychologist, 51(2), 247-265.

Klemmer, S. R., Hartmann, B., & Takayama, L. (2006, June). How bodies matter: five themes for interaction design. In Proceedings of the 6th conference on Designing Interactive systems (pp. 140-149).

Melcer, E. F., & Isbister, K. (2016, May). Bridging the physical divide: a design framework for embodied learning games and simulations. In Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems (pp. 2225-2233).

Micciolo, M., Arroyo, I., Harrison, A., Hulse, T., & Ottmar, E. (2018). The Wearable Learning Cloud Platform for the creation of embodied multiplayer math games. In Artificial Intelligence in Education: 19th International Conference, AIED 2018. Proceedings, Part II (pp. 220-224). Springer Intern. Publishing.

Nathan, M. J. (2021). Foundations of embodied learning: A paradigm for education. Routledge.

Nemirovsky, R., Kelton, M. L., & Rhodehamel, B. (2012). Gesture and imagination: On the constitution and uses of phantasms. Gesture, 12(2), 130–165. https://doi.org/10.1075/gest.12.2.02nem

Ma, J. Y. (2017). Multi-party, whole-body interactions in mathematical activity. Cogn. Instr. 35(2), 141-164.

Rasul, I., Castro, F., & Arroyo, I. (2023a, May). Towards Embodied Wearable Intelligent Tutoring Systems. *In Proc. of Int. Conf. on Intelligent Tutoring Systems* (pp. 298-306). Cham: Springer Nature Switzerland.

Rasul, I., Crabtree, D., Castro, F., Poh, A., Gattupalli, S. S., Kathala, K.., & Arroyo, I. (2023b). WearableLearning: Developing Computational Thinking Through Modeling, Simulation, and Computational Problem Solving. In *Proceedings ICLS*, pp. 1090-1093. International Society of the Learning Sciences.

Smith, H., Closser, A. H., Ottmar, E., & Arroyo, I. (2020). Developing math knowledge and computational thinking through game play and design: A professional development program. *Contemporary issues in technology and teacher education*, *20*(4), 660-686.

International Society of the Learning Sciences

From Solo to Collaborative: Design Affordances in Multi-User Immersive VR Simulations for Transforming Science Learning

Michelle Lui, Sunnie Gong, Chloe Lok
michelle.lui@utoronto.ca, sunnie.gong@mail.utoronto.ca, chloe.lok@mail.utoronto.ca
University of Toronto

Abstract: Immersive virtual reality (IVR) shows promise for science learning by enabling visualization of microscopic phenomena and dynamic processes. However, research has focused predominantly on individual learning experiences, overlooking the social nature of learning. This study investigates the collaborative affordances of multi-user IVR for science education through participatory design, involving six teams of undergraduate students and educators who co-created IVR learning simulations. Findings revealed seven science learning affordances: microscopic visualization, spatial relationships, scale traversal, representational connections, perspective-taking, temporal phenomena, and procedural skills—which multi-user IVR amplified through embodied communication, role-taking that grounds collaborative discourse, and distributed perspectives that require coordination. Analysis of designs highlights how narrative structures establish group-worthy joint tasks, fostering six collaborative activity types. Variations in structuring collaborative processes emerged from play-level scripts with undifferentiated roles, and scene-level scripts with defined roles enhancing interdependence. Results provide design insights for leveraging IVR's spatial and embodied features in collaborative science learning.

Introduction

Immersive virtual reality (IVR) has emerged as a promising technology for supporting science learning, particularly for visualizing abstract phenomena that are difficult to observe in traditional classroom settings (Coban et al., 2022; Matovu et al., 2023). Recent research demonstrates that situating learners within dynamic scientific models in IVR simulations enables engagement with microscopic molecular structures and processes previously inaccessible through direct observation (Lui et al., 2023; Ma et al., 2025; Reen et al., 2025). However, studies have focused primarily on individual learning experiences, overlooking a fundamental theoretical insight: learning occurs within social contexts (Stahl et al., 2022). While the effectiveness of computer-supported collaborative learning (CSCL) in STEM education is well-established (Jeong et al., 2019), current knowledge about collaborative learning in IVR remains limited (Lui et al., 2026; Paulsen et al., 2024; van der Meer et al., 2023). Without understanding how to design for effective collaboration in multi-user IVR, we risk perpetuating isolated learning in an inherently social medium.

Unlike screen-based collaborative environments, IVR offers co-presence, where learners share a three-dimensional space, use embodied interaction and spatial deixis naturally, engage in multimodal communication, and coordinate their attention across dynamic simulations. These characteristics suggest that collaborative IVR may support different forms of joint activity and distributed cognition. Moreover, current approaches to technology-rich educational design typically position learners as end users rather than as design partners. Participatory design methods center learners' and educators' insights throughout the process (DiSalvo et al., 2017), surfacing tacit knowledge about what supports effective collaboration. This study addresses the following:

RQ1: What are the design affordances for learning science with IVR?

RQ2: What are the design affordances of collaborative learning with scientific IVR simulations?

Background and theoretical framework

Prior reviews identify key rationales for adopting IVR in science education, including visualizing abstract concepts, experiential learning, and practical skills development, with visualization as the most prevalent (Matovu et al., 2023). IVR enables learners to explore spatial relationships in molecular structures, traverse scales from microscopic to macroscopic perspectives, and experience temporal phenomena that unfold too quickly or slowly to observe in real-time (Coban et al., 2022; Matovu et al., 2023)—supporting embodied cognition through natural gestures and movements (Johnson-Glenberg, 2018). However, while existing IVR applications incorporate sensory and actional design features, they often lack narrative and, critically, social features—reflecting a predominant focus on individual learning. Recent reviews have expanded our understanding of how IVR can support collaborative learning, identifying key affordances such as enhancing engagement, fostering social skill

 © ISLS

development, and enabling remote collaboration (van der Meer et al., 2023). However, they also note that many designs underutilize the collaborative potential of immersive environments (Hatch & Michaelis, 2024) and rely predominantly on play- and scene-level CSCL scripts (Lui et al., 2026). This study draws on Jeong and Hmelo-Silver's (2016) seven core CSCL affordances as its analytical lens: (1) engage in a joint task, (2) communicate, (3) share resources, (4) engage in productive processes, (5) engage in co-construction, (6) monitor and regulate collaborative learning, and (7) find and build groups and communities. To structure productive engagement, technology can deploy CSCL scripts—digital supports that guide learners engaged in collaborative tasks (Fischer et al., 2013). Vogel et al. (2017) described three different CSCL script levels of increasing socio-cognitive scaffolding complexity: play-level scripts involve undifferentiated roles, scene-level scripts involve role differentiation, while scriptlet-level scaffolds provide meta-cognitive supports to fulfill specific roles and responsibilities (Vogel et al., 2017).

Methods

Participants

Eighteen individuals (7 male, 11 female) from undergraduate life sciences programs at the University of Toronto participated: 14 students and 4 educators. Participants were organized into six teams (2-4 members each) of either mixed student-educator groups (4 teams) or student-only groups (2 teams). Most had minimal IVR experience. Table 1 shows team compositions and design focuses.

Procedure

Each design team participated in a series of three face-to-face or hybrid co-design sessions over a period of 6 weeks to 4 months. Three-hour sessions were facilitated by three researchers, using sticky notes or Miro (a web-based collaborative whiteboard platform) to capture and discuss participants' ideas.

Data sources and analysis

For RQ1, we analyzed written artifacts and video recordings from Session 1 discussions of challenging science concepts suitable for IVR. All video recordings were transcribed using text-to-speech software (Adobe Premiere Pro) and cleaned. Open codes from NVIVO 15 were transferred to Miro and examined inductively through affinity diagramming. For RQ2, finalized designs were converted into conjecture maps (Sandoval, 2014) and examined for the seven CSCL affordances (Jeong & Hmelo-Silver, 2016) and the science learning affordances identified in RQ1. The research team met weekly to discuss and stabilize interpretations, reviewing emerging codes collaboratively to build consensus and resolve discrepancies.

Table 1
Team Designs, Compositions, Pedagogical Goals, and Key Science Learning and CSCL Affordances Considered

Team (Code)	Composition	Pedagogical Goal	Science Learning*	CSCL**
CRISPR-Cas Design Simulator (CR)	1 educator, 2 students	Design CRISPR-Cas systems to treat genetic disease	1, 3	JT, Com, PP (scene)
Virtual qPCR Lab (QP)	1 educator, 2 students	Apply qPCR knowledge; develop lab fluency	4, 7	JT, Com, PP (play), CC
Imminent (IM)	1 educator, 2 students	Understand immune cell function and pathogen interactions	1, 5, 6	JT, Com, PP (scene), M&R, CB
Immune Blaster (IB)	4 students	Understand complex immunological response	1, 2, 3, 5, 6	JT, Com, PP (scene), M&R
Parkinson's Disease Patient Simulator (PD)	3 students	Practice clinical communication and team decision-making	5, 6	JT, Com, SR, PP (play), CC
Lion Chase (LC)	1 educator, 1 student	Visualize cortisol synthesis pathways across macro-micro scales	2, 3, 5, 6	JT, Com, PP (scene)

*Science Learning Affordances: 1=Microscopic phenomena, 2=Spatial relationships, 3=Traversing scales, 4=Representational connections, 5=Perspective-taking, 6=Temporal phenomena, 7=Procedural skills
**CSCL Affordances: JT=Joint task, Com=Communication, SR=Shared resources, PP=Productive processes (play/scene level), CC=Co-construction, M&R=Monitor & regulate, CB=Community building

 © ISLS

Findings

RQ1: Seven affordances for learning science with IVR

Analysis of participants' discussions revealed seven interconnected affordances: three address visualizing challenges, three bridge connections across representations and perspectives, and one supports embodied practice.

Visualization-scale in practice

Viewing microscopic phenomena. Eleven participants identified IVR's capacity to situate learners within microscopic environments, making molecular-level processes visible that are inherently 3D yet typically encountered through 2D abstract representations.

Visualizing spatial relationships. Fourteen participants noted IVR's value for learning anatomy through 3D models that can be viewed from multiple angles, resolving the depth-cue limitations of 2D representations.

Traversing scales. Seven participants emphasized IVR's ability to move back and forth across micro and macro levels, allowing learners to contextualize microscopic processes within macroscopic effects—for example, how molecular-level hormones affect organismal physiology.

Representation-perspective cluster in practice

Connecting representational forms. Three participants identified IVR's potential for helping learners make sense of mathematical representations by visualizing the corresponding molecular processes—bridging visual and mathematical forms of scientific representation.

Practicing perspective-taking. Five participants described how learners could embody biological entities (e.g., a cell performing its function) or adopt clinical perspectives (e.g., experiencing symptoms of a neurological condition), supported by IVR's immersive and embodied nature.

Experiencing temporal phenomena. Thirteen participants identified two modes: allocentric (3rd person observation of processes unfolding over time) and egocentric (1st person participation in biological processes).

Embodied practice

Developing procedural skills. Six participants identified IVR as valuable for learning laboratory skills, with embodied interactions more closely mirroring physical lab work than mouse-and-keyboard interactions.

Science learning affordances in action: The CRISPR-Cas Design Simulator (CR)

The CRISPR-Cas Design Simulator (CR) simulation exemplifies how viewing microscopic phenomena, visualizing spatial relationships, and scale traversal operate synergistically. Players in groups of three design a treatment for cystic fibrosis using CRISPR-Cas gene-editing technology. Players are shrunk to subcellular size and enter a cell nucleus, making genome structure and associated processes visible (viewing microscopic phenomena). A cooperative mini-game distributes views of the genome across three organizational scales (one per player): 3D genome structure, epigenetic interactions with histone proteins, and linear nucleotide sequences. To locate the target PAM sequence (by traversing scales and visualizing spatial relationships), coordination among players is required. Once the PAM sequence is identified, players collaborate to design the guide RNA and to determine the desired editing function of the CRISPR-Cas system. This transforms individual science learning IVR affordances (viewing microscopic phenomena, traversing scales) into meaningful collaboration that requires joint problem-solving.

RQ2: Collaborative affordances in multi-user IVR designs

Finding 1: Narrative structures enable diverse collaborative activities through joint tasks

All six designs established joint tasks through compelling narrative structures that created shared goals and authentic contexts. From competitive gameplay (IM, IB), clinical scenarios (PD), laboratory work (QP), to designing genetic therapies (CR). Within these narratives, six distinct types of collaborative activities emerged: (1) discussing, (2) consensus-building, (3) troubleshooting (identifying and resolving problems collaboratively), (4) strategizing, (5) designing, and (6) coordinating interdependent tasks. Competitive game narratives (IM, IB) naturally generated strategizing and coordination of interdependent tasks as teams planned attacks and defenses. Clinical scenarios (PD) required discussing observations and consensus-building for treatment plans. Laboratory simulations (QP) incorporated troubleshooting through error alerts and consensus-building through conceptual checkpoints. These narrative structures transform individual IVR affordances into collaborative necessities. For example, viewing microscopic phenomena becomes a shared perceptual experience that requires joint interpretation, and traversing scales becomes distributed among team members, requiring coordination.

 © ISLS

Finding 2: Role differentiation creates interdependence through scene-level scripts

Analysis through Vogel et al.'s (2017) three-level CSCL script framework revealed a continuum of collaborative structuring approaches. Two designs incorporated play-level scripts (PD, QP), in which learners oscillate between individual and collective roles without role differentiation. In the Parkinson's Disease Patient Simulator (PD) simulation, players engage with the same AI patient individually before being prompted to discuss and design a treatment plan for them collaboratively. Four teams utilized scene-level scripts (CR, IB, IM, LC), with complementary role differentiation that structured interdependent tasks. In CR, each player has a distinct main task while others offer support, with differing perceptual access to the genome during the mini-game. In IB, the Doctor/Drug and Patient/Immune System roles collaborate against the Virus role, each with distinct abilities and information. In IM, immune cell team members have differentiated biological functions that contribute to unique actions. Notably, all designs lacked scriptlet-level scaffolds—specific prompts supporting the metacognitive processes needed to fulfill assigned roles.

Finding 3: IVR's embodied nature enables unique forms of collaborative meaning-making

Beyond standard CSCL affordances, our analysis revealed IVR-specific collaborative features that extend the affordances for communication, joint tasks, and co-construction.

Embodied communication through gesture and spatial positioning. All six designs implicitly assumed that learners would use spatial positioning, gestures, and shared gaze—forms of non-verbal communication that mirror face-to-face collaboration. When learners point at shared virtual objects or position themselves relative to structures to indicate perspective, they engage in deictic communication that supports joint reference and shared attention, extending the communication affordance into 3D space.

Embodied role-taking for shared perspective understanding. Multiple designs (IM, IB, LC) incorporated embodied role-taking that goes beyond traditional role-playing by providing first-person sensory experiences that become objects of collaborative reflection. In IM, learners embody different biological entities—some players become immune cells (macrophages, dendritic cells, neutrophils) or *Salmonella* bacteria—with each role constraining actions according to biological functions, creating a foundation for collaborative understanding. In LC, learners experience simulated stress responses (e.g., tunnel vision through field-of-view blurring), creating shared visceral experiences that ground subsequent discussions about cortisol's effects.

Distributed perspectives requiring coordination for knowledge co-construction. Three designs (CR, LC, IB) strategically distributed perceptual access or abilities across team members, requiring coordination to achieve complete understanding. In the CR, team members simultaneously view the genome at different scales with unique possible actions. Locating the PAM sequence required communicating across perspectives and synchronizing actions.

Discussion

This study investigated design opportunities and collaborative affordances of multi-user IVR simulations for science learning through participatory design with undergraduates and educators. Our analysis of RQ1 identified seven key IVR affordances for science learning that extend Matovu et al.'s (2023) rationales for IVR adoption, emphasizing the interconnected nature of visualization, scale, and representation in scientific understanding. Importantly, our findings demonstrate how collaboration amplifies these affordances to leverage collective cognitive resources, transforming individual affordances into opportunities for distributed cognition and joint problem-solving (Paulsen et al., 2024; van der Meer et al., 2023). For RQ2, all six designs used narrative structures to establish joint tasks and shared goals with six collaborative activity types. Both the current study and Hatch & Michaelis (2024) identify discussion and problem-solving as essential collaborative processes; this study adds consensus-building, strategizing, designing, and coordinating interdependent tasks. Using Vogel et al.'s (2017) framework, two designs (PD, QP) employed play-level scripts, whereas four designs (CR, LC, IB, IM) utilized scene-level scripts, consistent with recent findings in IVR collaboration (Lui et al., 2026). Future IVR environments may benefit from incorporating more explicit scaffolding to elicit explanation, argumentation, and reasoning—forms of dialogue known to enhance CSCL learning (Fischer et al., 2013). Our findings illuminate how IVR's embodied and spatial nature creates distinctive opportunities beyond traditional CSCL. Designs leverage co-presence, synchronous communication that mirrors co-located face-to-face settings, enabling joint reference through gesture and spatial positioning that is difficult in remote, screen-based collaboration. IVR enables novel forms of embodied role-taking and perspective-sharing—providing first-person sensory experiences that become objects of collaborative reflection, amplifying emotional resonance and enabling learners to inhabit others' perspectives more fully (Järvelä & Rose, 2022). Since the findings are based on participatory design rather than on implementation and evaluation, future research should systematically explore the collaborative science learning opportunities that IVR offers.

References

Coban, M., Bolat, Y. I., & Goksu, I. (2022). The potential of immersive virtual reality to enhance learning: A meta-analysis. *Educational Research Review*, *36*, 100452. https://doi.org/10.1016/j.edurev.2022.100452

DiSalvo, B., Yip, J., Bonsignore, E., & DiSalvo, C. (Eds.). (2017). *Participatory design for learning: Perspectives from practice and research*. Routledge. https://doi.org/10.4324/9781315630830

Fischer, F., Kollar, I., Stegmann, K., & Wecker, C. (2013). Toward a script theory of guidance in computer-supported collaborative learning. *Educational Psychologist*, *48*(1), 56–66. https://doi.org/10.1080/00461520.2012.748005

Hatch, P., & Michaelis, J. E. (2024). A framework for collaborative learning in immersive virtual reality. *Proceedings of the ISLS Annual Meeting 2024*, 289–292. https://doi.org/10.22318/cscl2024.791972

Järvelä, S., & Rose, C. P. (2022). Novel CSCL design processes prompted by exploration of role-playing and emotions. *International Journal of Computer-Supported Collaborative Learning*, *17*(4), 457–461. https://doi.org/10.1007/s11412-022-09383-9

Jeong, H., & Hmelo-Silver, C. E. (2016). Seven affordances of computer-supported collaborative learning: How to support collaborative learning? How can technologies help? *Educational Psychologist*, *51*(2), 247–265. https://doi.org/10.1080/00461520.2016.1158654

Jeong, H., Hmelo-Silver, C. E., & Jo, K. (2019). Ten years of computer-supported collaborative learning: A meta-analysis of CSCL in STEM education during 2005–2014. *Educational Research Review*, *28*, 100284. https://doi.org/10.1016/j.edurev.2019.100284

Johnson-Glenberg, M. C. (2018). Immersive VR and education: Embodied design principles that include gesture and hand controls. *Frontiers in Robotics and AI*, *5*, 1–19. https://doi.org/10.3389/frobt.2018.00081

Kumar, S., Kaur, S., Seem, K., Kumar, S., & Mohapatra, T. (2021). Understanding 3D genome organization and its effect on transcriptional gene regulation under environmental stress in plant: A chromatin perspective. *Frontiers in Cell and Developmental Biology*, *9*. https://doi.org/10.3389/fcell.2021.774719

Lui, M., Chong, K.-Y. A., Mullally, M., & McEwen, R. (2023). Facilitated model-based reasoning in immersive virtual reality: Meaning-making and embodied interactions with dynamic processes. *International Journal of Computer-Supported Collaborative Learning*, *18*(2), 203–230. https://doi.org/10.1007/s11412-023-09396-y

Lui, M., Wang, Y., Yu, J., & Lok, C. (2026). Mapping design dimensions for collaborative learning in virtual reality: A scoping review. *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26)*. https://doi.org/10.1145/3772318.3791856

Ma, J., Cao, H., Nie, D., & Chen, D. (2025). Interactive molecular dynamics in virtual reality for multidisciplinary education: Theory and higher education applications. *Frontiers of Digital Education*, *2*(4), 36. https://doi.org/10.1007/s44366-025-0073-8

Matovu, H., Ungu, D. A. K., Won, M., Tsai, C.-C., Treagust, D. F., Mocerino, M., & Tasker, R. (2023). Immersive virtual reality for science learning: Design, implementation, and evaluation. *Studies in Science Education*, *59*(2), 205–244. https://doi.org/10.1080/03057267.2022.2082680

Paulsen, L., Dau, S., & Davidsen, J. (2024). Designing for collaborative learning in immersive virtual reality: A systematic literature review. *Virtual Reality*, *28*(1), 63. https://doi.org/10.1007/s10055-024-00975-4

Reen, F. J., Jump, O., McEvoy, G., McSharry, B. P., Morgan, J., Murphy, D., O'Leary, N., O'Mahony, B., Scallan, M., Walsh, C., & Supple, B. (2025). Student informed development of virtual reality simulations for teaching and learning in the molecular sciences. *Journal of Biological Education*, *59*(4), 604–620. https://doi.org/10.1080/00219266.2024.2386250

Sandoval, W. (2014). Conjecture mapping: An approach to systematic educational design research. *Journal of the Learning Sciences*, *23*(1), 18–36. https://doi.org/10.1080/10508406.2013.778204

Stahl, G., Koschmann, T., & Suthers, D. (2022). Computer-supported collaborative learning. In R. K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (3rd ed., pp. 406–427). Cambridge University Press. https://doi.org/10.1017/9781108888295.025

van der Meer, N., van der Werf, V., Brinkman, W.-P., & Specht, M. (2023). Virtual reality and collaborative learning: A systematic literature review. *Frontiers in Virtual Reality*, *4*. https://doi.org/10.3389/frvir.2023.1159905

Vogel, F., Wecker, C., Kollar, I., & Fischer, F. (2017). Socio-cognitive scaffolding with computer-supported collaboration scripts: A meta-analysis. *Educational Psychology Review*, *29*(3), 477–511. https://doi.org/10.1007/s10648-016-9361-7

International Society of the Learning Sciences

Epistemic Authority and Interface: A Typological Framework of AI's Roles in Knowledge Building

Lydia Cao, Wakana Tsuji, Robert Huang, Carl Bereiter, Marlene Scardamalia, Ahmad Khanlari, Boris Streipe
ly.cao@utoronto.ca, wakana.gen@mail.utoronto.ca, rob.huang@utoronto.ca, carl.bereiter@utoronto.ca
marlene.scardamalia@utoronto.ca, a.khanlari@mail.utoronto.ca, boris.steipe@utoronto.ca
Ontario Institute for Studies in Education, University of Toronto

Bodong Chen, cbd@upenn.edu, University of Pennsylvania
Sebastian Simon, sebastian.simon@univ-lemans.fr, Le Mans University

Yee Yin Tan, Seng Chee Tan
yeeyin.tan@nie.edu.sg sengchee.tan@nie.edu.sg
National Institute of Education, Nanyang Technological University of Singapore

Stefano Cacciamani, s.cacciamani@univda.it, University of Valle d'Aosta

Eliane Chaker Metni, Elaine Khoo
eliane.chaker@utoronto.ca, elaine.khoo@utoronto.ca
University of Toronto Scarborough

Caroline Traverse, Thérèse Laferrière
caroline.traverse.1@ulaval.ca, therese.laferriere@fse.ulaval.ca
Université Laval

Shotaro Naganuma, naganuma.shotaro.062@m.kyushu-u.ac.jp, Kyushu University
Jun Oshima, joshima@inf.shizuoka.ac.jp, Shizuoka University
Thomas Underwood, tunderwd@hku.hk, University of Hong Kong

Adiel Benjamin, Yotam Hod
adiel.benjamin@edtech.haifa.ac.il, yotamhod24@gmail.com
University of Haifa

Chandan Dasgupta, c.dasgupta@utwente.nl, University of Twente
Richard Reeve, richard.reeve@queensu.ca, Queen's University

Abstract: Knowledge Building distinguishes itself from other educational approaches by engaging students directly in the means by which knowledge advances in the world. As artificial intelligence (AI) advances, fundamental questions arise about its implications for KB. This study extended a workshop at the 2024 International Society of the Learning Sciences (ISLS) annual meeting into a four-month dialogue on Knowledge Forum. Using a dialogical knowledge-making framework, 23 contributors from diverse disciplinary backgrounds and career stages collectively examined theoretical, pedagogical, design, ethical implications at the intersection of AI and KB. Diverse AI roles were discussed and organized into a typological framework along *interface type* and *epistemic authority levels*, producing six configurations: directive interaction, collaborative interaction, exploratory interaction, diagnostic environment, reflective environment, and experimental environment. Given the configurations of AI, we propose a design principle of making AI's epistemic position visible and enabling learner-driven navigation across AI roles.

Introduction

The recent advances of generative artificial intelligence (genAI) have renewed important epistemological questions around knowledge building for societies: What is knowledge and how is knowledge constructed, by individuals and collectively, in the context of AI models capable of human-like linguistic productions (Jones & Bergen, 2024). Despite a growing body of research examining AI's role in education and research (Celik et al., 2022; Zheng et al., 2024), significant gaps remain in our understanding of how it might reshape learning, education, and Knowledge Building/Knowledge Creation.

Knowledge Building (Scardamalia & Bereiter, 2006) offers a particularly valuable lens for the discussion of AI and education, as it has historically championed student epistemic agency that augment rather than diminish human intellectual capacity and agency. Knowledge Building (KB) positions students as active epistemic agents who search, review, critique, and build on authoritative knowledge sources to advance the state of their community

 © ISLS

knowledge rather than simply consume and regurgitate for standardized testing. However, our understanding of how AI functions within KB contexts remains nascent and evolving. While AI technologies advance rapidly, frameworks for conceptualizing AI's role in KB are still emerging.

Given the complexity and rapid evolution of this landscape, there is a pressing need to bring together a plurality of perspectives and experiences to collectively examine how AI intersects with KB to build frameworks that capture and meaningfully include AI in KB. Using a dialogical knowledge-making framework inspired by postdigital dialogue methodology (Jandrić et al., 2019; Markauskaite et al., 2022), the present study brought 23 scholars together across disciplines who share common interests in KB. The goal of this endeavor is to initiate a process of identifying theoretical, pedagogical, design, and ethical implications of AI in KB communities and to start mapping out the evolving relationship between AI and KB. This short paper reports on one research question that was explored during the collaborative dialogue: What roles can AI play for Knowledge Building?

Method

The study's design is informed by and built upon Markauskaite et al.'s (2022) cyclical reflection approach with post digital dialogue and experimental multi-authored methodology. Given the novelty of technology and a lack of studies addressing the theory of KB and AI, we chose a critical communicative methodology (Gomez et al., 2018), to include stakeholders and actors from different backgrounds in a continuous, dialogic, constructive approach, allowing for nuanced, co-constructed perspectives that traditional methods (such as surveys and interviews) cannot provide.

Contributors

This collaborative dialogue involved 23 contributors from 13 distinct disciplines, with many identifying across multiple fields (n>23 as participants could select multiple disciplines). The most represented discipline was Learning Sciences (n=15), followed by Computer-Supported Collaborative Learning (n=7), and Educational Technology (n=4). Other disciplines included Computer Science (n=2), and single representations in Psychology, Computational Biology, Language Education, Human-Computer Interaction, Learning Analytics, Developmental and Educational Psychology, and Vocational and Health Sciences Education.

Research context and procedure

The dialogic space for participants in this study emerged during a workshop titled "Knowledge building and generative AI to advance community knowledge", at the International Society of the Learning Sciences (ISLS) in Buffalo, US, June 2024 (Cao et al., 2024). This study reports on the dialogue continued from the workshop, a process that lasted from July to October 2024 and culminated in this collective paper. The collective dialogue took place through written exchanges in Knowledge Forum (KF), which allows participants to create notes on an infinite canvas for their ideas, on which other participants can then build on. The collaborative dialogue was organized into three phases. *Phase 1 Open Discussion:* Participants were invited to contribute to a shared space on Knowledge Forum in response to prompting questions after the workshop. After posting initial responses, participants engaged with at least three other contributors' KF notes. *Phase 2 Clustering of Notes*: In Knowledge Forum, the first and second author manually organized contributors' notes into clusters based on the convergence and divergence of ideas in ways to maximize the potential for further idea improvement and rise-above (Scardamalia, 2002). Participants were then encouraged to review and build on their "near-neighbors" and "far-neighbors" notes from different clusters. *Phase 3 Report Draft:* The first and second author synthesized the contributions from the two phases and the workshop to an initial report. The draft was shared with all the contributors for both member checking and refinement of ideas.

Data sources and analysis

Data of this study comes from the Knowledge Forum's notes (n=91) contributed by the 23 authors. The first and second authors engaged in Braun and Clarke's (2019) reflexive thematic coding in ATLAS.ti, a qualitative research method for identifying, analyzing, and reporting patterns within data after the three phases of collective dialogue. Using a reflexive inductive thematic analysis approach, researchers are positioned as an instrument central to generating rich, situated insights. Rather than treating themes and insights as pre-existing truths to be "discovered," they were iteratively constructed and refined through cycles of coding and dialogue with the data. This analytical approach is in alignment with the dialogic nature of knowledge making and the theoretical foundation of postdigital dialogue (Jandrić et al., 2019).

Findings

 © ISLS

In this paper, voices from the collaborative KF dialogue are attributed by contributor surname followed by "KF" (e.g., Metni, KF), while external scholarly sources follow standard APA citation format. Through the collaborative dialogue, a variety of AI roles emerged, which we organized into six configurations along two dimensions: 1) the interface between humans and AI (either conversational or embedded), and 2) the epistemic authority assumed by or attributed to AI. In conversational interfaces, humans explicitly communicate with AI through conversation (e.g., chatbots, virtual assistants). When AI is embedded in the environment, it functions as part of the KB environment itself. The epistemic authority dimension refers to the degree to which AI is positioned or assumes epistemic authority. According to Jäger (2025, p. 1), epistemic authority is the authority we ascribe to people "in virtue of their favorable relation to epistemic goods such as true belief, rational credence, knowledge, or understanding". It involves giving "special epistemic weight" to someone's views. Though large language models (LLMs) do not possess true expertise or knowledge in the epistemological sense that they don't have genuine understanding, beliefs, or direct access to reality (Bender & Koller, 2020; Mancoridis et al., 2025; Mitchell & Krakauer, 2023), AI systems are frequently anthropomorphized and positioned with varying degrees of epistemic authority in educational contexts (Jose et al., 2025). In this typology, we organized AI's epistemic authority into high, shared/equal, and low. Table 1 presents the six configurations arising from the interaction between *interface* and *epistemic authority*, which are further explained below.

Table 1.
Typological Framework of Six Configurations of AI's Roles in Knowledge Building

	Configuration	Interface	Epistemic Authority	Description	Illustrative Roles
1	**Directive Interaction**	Conversational	**High**	AI engages learners via conversation from a position of authority, providing direct instruction, expert knowledge, and authoritative feedback	*Tutors, experts, guides*
2	**Collaborative Interaction**	Conversational	**Shared**	AI engages learners via conversation as a co-constructive partner with shared epistemic authority, supporting psychological safety and joint knowledge construction	*Companions, partners, consultants, facilitators, brokers*
3	**Exploratory Interaction**	Conversational	**Low**	AI engages learners via conversation and introduces alternative voices and perspectives without claiming epistemic authority	*Group members, simulators, debate partners*
4	**Diagnostic Environment**	Embedded	**High**	AI is woven into the knowledge building environment and evaluates contributions against predetermined criteria, providing assessment and feedback	*Formative assessment, explanatory coherence feedback*
5	**Reflective Environment**	Embedded	**Shared**	AI is woven into the knowledge building environment and surfaces patterns and visualizations of community knowledge, requiring learner interpretation and collective decision-making	*Interactive and concurrent analytics, knowledge space visualization*
6	**Experimental Environment**	Embedded	**Low**	AI is woven into the knowledge building environment as a responsive, open-ended space where learners can play with, prototype, and refine ideas and language through experimentation	*Conceptual playgrounds, language scaffolds*

 © ISLS

Directive interaction (conversational, high epistemic authority)
In this configuration, AI engages learners via conversation from a position of authority, providing direct instruction, "expert" knowledge, and authoritative feedback. The AI's outputs are given great epistemic weight, withz its assertions treated as authoritative to an extent. This configuration reflects what Tan (2024) describes as "learning from AI". In other words, AI plays the role of a more knowledgeable other in this configuration, such as tutors, experts, and guide.

Collaborative interaction (conversational, shared epistemic authority)
In this configuration, AI engages learners via conversation as a co-constructive partner with shared epistemic authority, supporting psychological safety and joint knowledge construction, including roles as companion, partner, consultant, facilitator, or broker with shared epistemic authority with learners. This configuration aligns with Tan's (2024) notion of "learning with AI". Such positioning of AI as companion, partner, consultant, or broker, with shared epistemic authority differs from the directive interaction configuration, where AI assumes more authority.

Exploratory interaction (conversational, low epistemic authority)
In this configuration, AI engages learners via conversation and introduces alternative voices and perspectives without assuming epistemic authority by playing the roles of group members (Bereiter & Scardamalia, KF) or other characters/persona (Huang, Chen, KF).

Diagnostic environment (embedded, high epistemic authority)
In this configuration, AI are integrated directly into the KB environment (e.g. authoring functionality in software like Knowledge Forum) and are positioned with more epistemic authority to formatively assess, diagnose, and provide feedback on both individual contributions and community knowledge advancement.

Reflective environment (embedded, shared epistemic authority)
This configuration represents AI systems that are embedded into the KB environment with shared epistemic authority between humans and AI. Rather than providing assessments predefined by certain criteria, these systems create interactive spaces that require human interpretation and decision-making. The AI-enhanced interactive and concurrent analytics systems are designed to move beyond Scardamalia and Bereiter's (KF) critique of AI offering "summarization [that] provides a simple accounting of topics and issues in the discourse... tending to close discussion and leave participants with an impression of a job well done."

Experimental environment (embedded, low epistemic authority)
This configuration transcends the conventional notions of "assistance" or "partnership" to position AI as part of the epistemic infrastructure proposed by Chen (2025) — not as an authority or companion, but as a responsive space that augments human creative and conceptual work. Instead of asking what AI can do for KB, this approach asks how AI can expand what students can do with ideas. Building on the sandbox metaphor of knowledge creation, As Tsuji (KF) explains, "AI could be idea- and text-building space for personal (and perhaps also shared) reflection and experiment through remixing, prototyping, testing, and refining ideas and texts to support ideational writing in Knowledge Building."

Discussion, limitations, and future directions

The collaborative dialogue documented in this short paper reveals that the intersection of Knowledge Building and generative AI is not merely a matter of technological integration but represents a fundamental reconceptualization of what it means to engage in knowledge building in the AI age. One of the main contributions of this collaborative dialogue highlighted in this short paper is the typological framework that maps AI role by interface type (conversational vs. embedded) and epistemic authority levels (high, shared, low), creating various configurations. We thus propose a design principle that suggests that AI-powered KB environments should allow users to fluidly move between these roles based on pedagogical needs and objectives. Effective design should support learners in dynamically coordinating different AI roles for Knowledge Building. Rather than fixing AI in predetermined roles, effective design provides learners with agency to dynamically position AI along the typological framework as their knowledge work requires, and their capacity of productive human-AI collaboration develops. We acknowledge the limitations of our work, for instance the potential bias induced by the authors conducting the thematic analysis of data the production of which they were actively involved in. The derived insights further originate from a pool of participants with the majority being researchers with a learning science background. Future research should broaden these perspectives by incorporating voices from researchers with other backgrounds, teachers, administrators, policy-makers, and learners, and by analyzing cross-cultural contexts.

References

Bender, E. M., & Koller, A. (2020). Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 5185–5198. https://doi.org/10.18653/v1/2020.acl-main.463

Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health*, *11*(4), 589–597. https://doi.org/10.1080/2159676X.2019.1628806

Cao, L., Scardamalia, M., Bereiter, C., Costa, S., Resendes, M., Soliman, D., Khanlari, A., Steipe, B., Zhang, J., Chen, M.-H., Chen, B., Laferrière, T., Yang, Y., Chan, C. K. K., Zhu, G., Teo, C. L., Castillo, F., Raman, P., & Massey, L. (2024). Knowledge Building and Generative AI to Advance Community Knowledge. In C. Hoadley & C. Wang (Eds.), *General Proceedings of the 4th Annual Meeting of the International Society of the Learning Sciences 2024* (pp. 106–109). International Society of the Learning Sciences.

Celik, I., Dindar, M., Muukkonen, H., & Järvelä, S. (2022). The Promises and Challenges of Artificial Intelligence for Teachers: A Systematic Review of Research. *TechTrends*, *66*(4), 616–630. https://doi.org/10.1007/s11528-022-00715-y

Chen, B. (2025). *Beyond Tools: Generative AI as Epistemic Infrastructure in Education* (arXiv:2504.06928). arXiv. https://doi.org/10.48550/arXiv.2504.06928

Gomez, K., Kyza, E. A., & Mancevice, N. (2018). Participatory Design and the Learning Sciences. In *International Handbook of the Learning Sciences*. Routledge.

Jäger, C. (2025). Epistemic Authority. In J. Lackey & A. McGlynn (Eds.), *Oxford Handbook of Social Epistemology*. https://philarchive.org/rec/JGEEA

Jandrić, P., Ryberg, T., Knox, J., Lacković, N., Hayes, S., Suoranta, J., Smith, M., Steketee, A., Peters, M., & McLaren, P. (2019). Postdigital dialogue. *Postdigital Science and Education*, *1*(1), 163–189.

Jones, C. R., & Bergen, B. K. (2024). *People cannot distinguish GPT-4 from a human in a Turing test* (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2405.08007

Jose, B., Cleetus, A., Joseph, B., Joseph, L., Jose, B., & John, A. K. (2025). Epistemic authority and generative AI in learning spaces: Rethinking knowledge in the algorithmic age. *Frontiers in Education*, *10*. https://doi.org/10.3389/feduc.2025.1647687

Mancoridis, M., Weeks, B., Vafa, K., & Mullainathan, S. (2025). *Potemkin Understanding in Large Language Models* (arXiv:2506.21521). arXiv. https://doi.org/10.48550/arXiv.2506.21521

Markauskaite, L., Marrone, R., Poquet, O., Knight, S., Martinez-Maldonado, R., Howard, S., Tondeur, J., De Laat, M., Buckingham Shum, S., Gašević, D., & Siemens, G. (2022). Rethinking the entwinement between artificial intelligence and human learning: What capabilities do learners need for a world with AI? *Computers and Education: Artificial Intelligence*, *3*, 100056. https://doi.org/10.1016/j.caeai.2022.100056

Mitchell, M., & Krakauer, D. C. (2023). The debate over understanding in AI's large language models. *Proceedings of the National Academy of Sciences*, *120*(13), e2215907120. https://doi.org/10.1073/pnas.2215907120

Scardamalia, M. (2002). Collective cognitive responsibility for the advancement of knowledge. In B. Smith (Ed.), *Liberal education in a knowledge society* (pp. 67–98). Open Court.

Scardamalia, M., & Bereiter, C. (2006). Knowledge building: Theory, pedagogy, and technology. In K. Sawyer (Ed.), *Cambridge Handbook of the Learning Sciences* (pp. 97–118). Cambridge University Press.

Tan, S. C. (2024). *Learning from, learning with, learning about, and learning beyond AI*. https://dr.ntu.edu.sg/handle/10356/181097

Zheng, Y., Li, X., Huang, Y., Liang, Q., Guo, T., Hou, M., Gao, B., Tian, M., Liu, Z., & Luo, W. (2024). Automatic Lesson Plan Generation via Large Language Models with Self-critique Prompting. In A. M. Olney, I.-A. Chounta, Z. Liu, O. C. Santos, & I. I. Bittencourt (Eds.), *Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners, Doctoral Consortium and Blue Sky* (pp. 163–178). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-64315-6_13

International Society of the Learning Sciences

Modeling and Analyzing Collaborative Dialogue States in Mathematics Problem-Solving through ONA and Hidden Markov Models

Shan Zhang, University of Florida, zhangshan@ufl.edu
Andrés Felipe Zambrano, University of Pennsylvania, azamb13@upenn.edu
Hongming Li, University of Florida, hli3@ufl.edu
Seiyon Lee, University of Florida, leeseiyon@ufl.edu
Ji-Eun Lee, Singapore University of Technology and Design, jieun_lee@sutd.edu.sg
Anthony F. Botelho, University of Florida, abotelho@coe.ufl.edu

Abstract: Prior research has shown that collaborative dialogue reflects students' cognitive, social, and regulatory processes, which support deeper learning. However, less is known about how these moves unfold sequentially during collaborative problem solving (CPS), especially in math contexts. This study leverages Markov Chains, Hidden Markov Models (HMM), and Ordered Network Analysis (ONA) to capture and analyze recurring patterns and transitions in dialogue that are representative of collaboration dynamics and observe how these relate to students' perceptions of their interactions, collaboration quality, and performance. We conduct our analyses with data from 22 college students (11 dyads) who worked collaboratively to solve algebra problems in a digital learning environment. Results show that dialogue within problems cycled through agreement, clarification, and self-explanation, indicating stable interaction patterns across five HMM latent states. Findings from ONA suggest that while explanatory talk anchors collaboration, task-grounding and co-constructive exchanges may be more directly associated with effective CPS.

Introduction

A student's capacity to express their mathematical reasoning through oral and written communication is considered essential to demonstrating mastery (Partnership for 21st Century Skills, 2007). One effective way to foster such communication is by engaging students in *Collaborative Problem Solving* (CPS), an instructional approach in which groups of learners work together to solve novel problems by sharing their understanding and coordinating their efforts (Graesser et al., 2020; Andrews-Todd & Forsyth, 2023). As CPS involves students engaging in sustained, multimodal interaction (e.g., combinations of verbal and non-verbal exchanges to achieve a shared goal), this complex phenomenon poses challenges for researchers seeking to capture and analyze patterns of collaboration and discourse quality. Prior studies have often examined dialogue in isolation (Tausczik et al., 2014), focusing on individual communicative moves such as building, monitoring, or repairing shared understanding (e.g., Hansen, 2022), rather than exploring how these acts unfold sequentially to shape the dynamics of collaborative reasoning. Understanding how these dialogue acts co-occur and form sequences or latent states is essential, as such patterns reveal the dynamic processes through which students shift between productive and unproductive forms of collaboration (Earle-Randell et al., 2023). With advancements in Computer-supported Collaborative Learning (CSCL) tools and data analytics, researchers have begun applying advanced modeling techniques, such as Hidden Markov Models (HMM) (Earle-Randell et al., 2023) or Ordered Network Analysis (ONA), to capture these temporal and structural dynamics more precisely (Tan et al., 2023). Yet, despite these methodological developments, relatively few studies have applied such approaches within mathematics education, where the cognitive and representational complexity of problem solving demands fine-grained analysis. As a result, considerable uncertainty remains about how best to unpack the intricate interplay of dialogue, reasoning, and regulation that characterizes CPS in this context.

To address these gaps, the present study examines the dynamic structure of collaborative dialogue during mathematical problem solving, focusing on how latent interaction states relate to collaboration quality, prior experience, perception of interaction, and learning outcomes. Specifically, it investigates the following research questions: RQ1: (a). How do observable dialogue acts transition from one to another during collaboration? (b). What latent problem-solving states emerge from students' dialogue during collaborative tasks, and how do students transition among these hidden states? RQ2: To what extent are HMM state proportions associated with the outcomes (i.e., overall learning satisfaction, perceptions of collaboration, problem-solving correctness, collaboration quality) as well as prior math learning collaborative experience and prior collaborative skills? RQ3:

 © ISLS

How do patterns of collaborative dialogue acts differ between groups with high versus low quality of collaborative problem-solving?

Methods

This study involved 22 undergraduate students (11 dyads) from a university located in the northeastern U.S. Of these students, 55% (n = 12) were male and 45% (n = 10) female. In terms of race/ethnicity, 64% (n = 14) identified as White, 32% as Asian (n = 5), and the remaining students (n = 3) as other categories. In this study, we collected data through Graspable Math (GM, https://activities.graspablemath.com/), a dynamic mathematics notation platform grounded in perceptual learning and embodied cognition theories. Prior to the task, students completed a pre-survey assessing collaborative experience and skills, followed by a brief practice phase. Dyads then worked together for 30 minutes to solve five algebra problems, after which they completed a post-survey measuring collaboration perceptions and satisfaction. Group-level problem-solving correctness was also computed. Multimodal data, including dialogue transcripts and interaction logs, were collected and temporally aligned to capture collaborative problem-solving processes. In total, 11 groups completed 49 problems, and all analyses were conducted at the group level.

To capture interaction processes, dialogue transcripts were coded using an established scheme (Zhang et al., 2024) encompassing cognitive, metacognitive, social, and affective dimensions ($\kappa > .75$). In parallel, collaboration quality was evaluated using a six-dimension rubric on a three-point scale, achieving high inter-rater reliability (QWK = .84). Each group's overall score was calculated by averaging its total problem scores. Finally, we used the median of these group-level overall scores to create two categories: high-quality collaboration (scores greater than or equal to the median; N = 6) and low-quality collaboration (scores below the median; N = 5). Detailed descriptions of the learning environment, coding scheme, and rubric are available on Open Science Framework: https://osf.io/583g2/overview?view_only=3120f65ebca541378e5c9d6ce00db05e. To analyze collaboration dynamics, we first applied first-order Markov chains to model transitions between dialogue acts (Thomas et al., 1983), followed by a Hidden Markov Model (Boyer et al., 2011) with five latent states to identify underlying interaction patterns from time-ordered sequences, with model selection based on Bayesian Information Criterion with five-fold cross-validation. We then examined emission probabilities, transition patterns, and run-lengths, and derived group-level features such as state proportions (time spent in each state). Next, to examine associations with outcomes, we computed Spearman correlations between hidden-state proportions and measures of collaboration perceptions, satisfaction, correctness, collaboration quality, and prior experience and skills. Finally, we conducted Ordered Network Analysis using the ONA Web Tool (version 1.7.0; Marquart et al., 2018) at the group level to compare structural dialogue patterns between high- and low-collaboration groups (median split), retaining key dialogue acts based on connection weights (CW > .02) and testing group differences using Mann-Whitney U on unit means after means rotation.

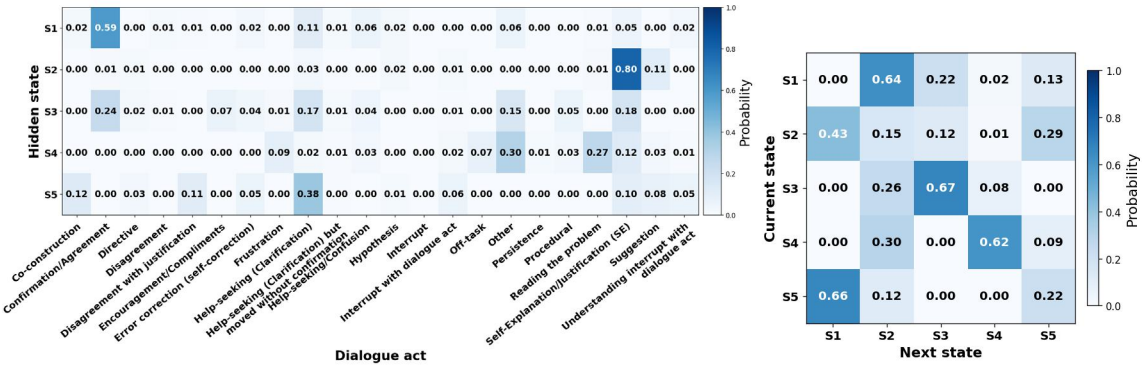
Results

In addressing RQ1, the Markov chain analysis revealed systematic and asymmetric transition patterns among dialogue acts. Confirmation/Agreement functioned as a central hub, frequently following Directives (P = 0.42), Help-seeking (P = 0.31), and Self-Explanation/Justification (SE) (P = 0.24). In turn, SE was often followed by Confirmation/Agreement (P = 0.49) or Help-seeking (P = 0.33), suggesting that confirmation and clarification prompted further reasoning. Help-seeking also commonly emerged after Disagreement with justification (P = 0.20), Hypothesis (P = 0.28), and Suggestions (P = 0.15), supporting the resolution of uncertainty. In contrast, disagreement-related moves were typically short-lived and often transitioned into Agreement (P = 0.14-0.37), indicating a tendency to quickly re-establish consensus. Less frequent acts (e.g., Frustration, Persistence, Off-task) appeared sporadically and showed limited continuation. Overall, collaboration was characterized by recurring cycles of agreement, help-seeking, and self-explanation.

Figure 1

HMM Emission Means (Left; N = 5) and Group-level HMM Transition Probabilities (Right)

 © ISLS



HMM analysis identified five latent states underlying CPS (Figure 1). State 1 (run length = 1) reflected brief agreement-dominant exchanges, while State 2 (run length = 1) captured frequent self-explanation and suggestion (SE: P = 0.80). State 3 (run length = 4) represented more sustained, mixed interaction combining agreement, clarification, and partial explanation. State 4 (run length = 3) reflected task grounding (e.g., reading, filler talk, occasional frustration), and State 5 (run length = 1) captured help-seeking, co-construction, and justified disagreement. Transition patterns showed strong cycling between States 1 and 2, shifts from State 2 to State 5, and high self-transitions in States 3 and 4, indicating more stable interaction phases.

To address RQ2, as shown in Table 1, the correlation analysis revealed several connections between hidden-state proportions and collaborative-related pre- or post-outcomes. State 4 (Reading the problem/Other) showed a moderate positive correlation with overall collaboration perceptions and a weaker, non-significant trend with satisfaction. State 5 (*Help-seeking/Co-construction/Disagreement with Justification*) was positively associated with group-level collaboration quality, suggesting that students spending more time in this state perceived their groups as more collaborative. State 2 (*SE-dominant*) was strongly associated with prior experience in math collaboration, indicating that students with richer backgrounds were more likely to engage in *Self-Explanation/Justification*. By contrast, State 3 (*Mixed*) consistently showed negative, non-significant correlations across outcomes. Overall, only States 4, 5, and 2 showed statistically significant associations with collaboration outcomes or prior experience.

Table 1

Spearman Correlations between Hidden-state Proportions (S1–S5) and Collaborative Measures

	S1	S2	S3	S4	S5
Collaborative learning Satisfaction	0.19	0.18	-0.21	0.4	0.14
Overall perceptions of collaboration	0.29	0.15	-0.38	0.43*	0.12
Overall problem-solving correctness	0.12	0.05	0.27	-0.03	-0.04
Group-level collaboration quality	0.18	0.27	-0.24	0.09	0.46*
Prior math learning collaboration experience	-0.04	0.53**	-0.36	-0.14	0.11
Prior collaborative skills	0.22	0.11	-0.17	-0.12	-0.2

*Note. *p < .05, **p < .01, ***p < .001.*

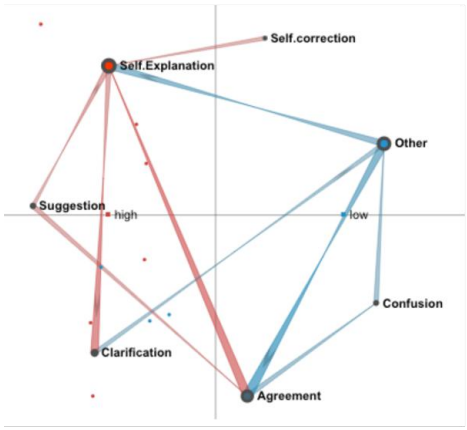
In addressing RQ3, Figure 2 presents the ONA difference model comparing high-collaboration (red, *N* = 6) and low-collaboration (blue, *N* = 5) groups. A Mann–Whitney U test confirmed significant differences in overall dialogue patterns (U = 27.00, *p* = .03). High-collaboration groups engaged more actively in self-explanation (SE; CW = 0.583 vs. 0.468) and showed stronger bidirectional connections between explanation and clarification (CW = 0.277 and 0.263 vs. 0.232 and 0.230). These groups were also more likely to transition from explanation to agreement (CW = 0.373 vs. 0.322), error correction (CW = 0.054 vs. 0.032), and suggestion (CW = 0.081 vs. 0.061), forming sustained cycles of co-construction. In contrast, low-collaboration groups produced more "Other" utterances (CW = 0.139 vs. 0.039), reflecting filler talk, and more often surrounded explanations with such contributions. Their agreements were more frequently preceded by non-substantive utterances (CW = 0.131 vs. 0.073), suggesting shallower consensus. They also showed higher tendencies toward confusion (CW =

 © ISLS

0.035 vs. 0.007) and clarification without deeper elaboration (CW = 0.079 vs. 0.055), further distinguishing their interaction patterns from high-collaboration groups.

Figure 2

ONA Difference Model of High (Red) vs. Low (Blue) Quality Groups



Connection Weights of ONA Comparing High and Low Collaboration Groups.

Connection	High Collab.	Low Collab.	Difference
Self-Explanation → Self-Explanation	0.583	0.468	0.115
Other → Other	0.035	0.139	-0.104
Other → Agreement	0.073	0.131	-0.059
Self-Explanation → Agreement	0.373	0.322	0.051
Agreement → Other	0.073	0.120	-0.048
Self-Explanation → Clarification	0.277	0.230	0.047
Other → Self-Explanation	0.097	0.139	-0.042
Self-Explanation → Other	0.122	0.159	-0.038
Suggestion → Self-Explanation	0.097	0.066	0.031
Clarification → Self-Explanation	0.263	0.232	0.030
Confusion → Agreement	0.019	0.048	-0.029
Other → Confusion	0.007	0.035	-0.028
Agreement → Suggestion	0.062	0.036	0.027
Clarification → Other	0.056	0.082	-0.026
Other → Clarification	0.055	0.079	-0.025
Self-Explanation → Self-correction	0.054	0.032	0.023
Self-Explanation → Suggestion	0.081	0.061	0.020

Discussion and implications

This study combined HMM, correlational analyses, and ONA to examine how dialogue patterns shape students' experiences and outcomes in CPS. Across methods, results highlight that productive collaboration is driven by multiple, interacting dialogue processes rather than a single dominant behavior.

The HMM revealed how structural patterns of dialogue acts distinguish collaboration quality by uncovering latent states and their transitions. For example, State 2 (SE-dominant) emerged as the dominant hub, occupying the largest share of time and mediating transitions across states. This state aligns with Mercer's category of *exploratory talk* (2000), which involves making reasoning explicit and ensuring mutual understanding. Prior work showed that exploratory talk dominates collaborative dialogue and sustains the longest dwell times (Earle-Randell et al., 2023). While exploratory talk more broadly includes justification, in this study, we found that questioning, elaboration, self-explanation, and clarification also function as its core markers. Similarly, the ONA results revealed self-explanation to be central to high-quality collaboration, aligning with prior work exploring mediators of cognitive challenge (Liang & Fung, 2020; Patterson, 2018).

The remaining states played complementary roles. State 1 captured agreement-dominant exchanges characteristic of *cumulative talk* (e.g., uncritical agreement, repetition, and passive acceptance), which may preserve social harmony but contribute little to deeper understanding (Hennessy et al., 2016; Mercer, 1995, 2000; Mercer & Wegerif, 1999). State 3 reflected mixed dialogue, State 4 focused on task grounding through reading and reframing, and State 5 combined help-seeking with co-construction and disagreement with justification. Correlational analyses further revealed that State 4 was positively associated with students' perceptions of collaboration, suggesting that grounding activities strengthen shared understanding, whereas State 5 correlated with collaboration quality, indicating that help-seeking and co-construction, especially when paired with justified disagreement, act to indicate sustained productive interaction. State 2 was linked to students' prior collaborative experience in mathematics, implying that explanatory and clarifying dialogue reflects transferable strategies from earlier contexts.

Unlike HMM, which captures temporal transitions among latent dialogue states, the ONA comparison highlighted structural differences between high- and low-quality groups. High-quality groups exhibited recurring cycles of self-explanation, clarification, agreement, and suggestion, consistent with Mercer's framework of exploratory talk, in which students justify reasoning, co-construct knowledge, and critically engage with peers' ideas. In contrast, low-quality groups relied more heavily on filler Other utterances and expressions of confusion, often surrounding explanations with non-substantive talk, aligning with Earle-Randell et al.'s (2023) finding that confusion can lead to disputational talk and conflict. These findings suggest that while *explanatory talk* (State 2) anchors collaboration, it is not sufficient alone. Task-oriented grounding (State 4) enhances shared perceptions of collaboration, and problem-focused exchanges involving help-seeking, co-construction, and justified disagreement (State 5) directly support collaboration quality. Overall, our results indicate that high-quality collaboration is best characterized not by a single dominant dialogue move but by the interplay of explanatory, grounding, and co-constructive processes.

References

Andrews-Todd, J., & Forsyth, C. M. (2023). Assessment of collaborative problem solving skills. In *International Encyclopedia of Education(Fourth Edition)* (pp. 494–503). Elsevier. https://doi.org/10.1016/B978-0-12-818630-5.10062-4

Boyer, K. E., Phillips, R., Ingram, A., Ha, E. Y., Wallis, M., Vouk, M., & Lester, J. (2011). Investigating the relationship between dialogue structure and tutoring effectiveness: a hidden Markov modeling approach. *International Journal of Artificial Intelligence in Education*, *21*(1-2), 65-81.

Earle-Randell, T. V., Wiggins, J. B., Ruiz, J. M., Celepkolu, M., Boyer, K. E., Lynch, C. F., Israel, M., & Wiebe, E. (2023). Confusion, Conflict, Consensus: Modeling Dialogue Processes During Collaborative Learning with Hidden Markov Models. In N. Wang, G. Rebolledo-Mendez, N. Matsuda, O. C. Santos, & V. Dimitrova (Eds.), *Artificial Intelligence in Education* (Vol. 13916, pp. 615–626). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-36272-9_50

Graesser, A. C., Greiff, S., Stadler, M., & Shubeck, K. T. (2020). Collaboration in the 21st century: The theory, assessment, and teaching of collaborative problem solving. *Computers in Human Behavior*, *104*, 106134. https://doi.org/10.1016/j.chb.2019.09.010

Hansen, E. K. S. (2022). Students' agency, creative reasoning, and collaboration in mathematical problem solving. *Mathematics Education Research Journal*, *34*(4), 813–834. https://doi.org/10.1007/s13394-021-00365-y

Hennessy, S., Rojas-Drummond, S., Higham, R., Márquez, A. M., Maine, F., Ríos, R. M., García-Carrión, R., Torreblanca, O., & Barrera, M. J. (2016). Developing a coding scheme for analysing classroom dialogue across educational contexts. *Learning, Culture and Social Interaction*, *9*, 16–44. https://doi.org/10.1016/j.lcsi.2015.12.001

Liang, W., & Fung, D. (2020). Development and evaluation of a WebQuest-based teaching programme: Students' use of exploratory talk to exercise critical thinking. *International Journal of Educational Research*, *104*, 101652. https://doi.org/10.1016/j.ijer.2020.101652

Marquart, C. L., Hinojosa, C., Swiecki, Z., Eagan, B., & Shaffer, D. W. (2021). Epistemic network analysis (Version 1.7.0) [Software]. Available from http://app.epistemicnetwork.org

Mercer, N. (1995). *The Guided Construction of Knowledge: Talk Amongst Teachers and Learners*. Multilingual Matters.

Mercer, N. (2000). *Words and Minds*. Routledge. https://doi.org/10.4324/9780203464984

Mercer, N., & Wegerif, R. (1999). Is 'exploratory talk' productive talk? In *Learning with Computers*. Routledge.

Patterson, E. W. (2018). Exploratory talk in the early years: Analysing exploratory talk in collaborative group activities involving younger learners. *Education 3-13*, *46*(3), 264–276. https://doi.org/10.1080/03004279.2016.1243141

Partnership for 21st Century Skills (2007). *Partnership for 21st century skills - Core content integration.* https://www.marietta.edu/sites/default/files/documents/21st_century_skills_standards_book_2.

Tan, Y., Ruis, A. R., Marquart, C., Cai, Z., Knowles, M. A., & Shaffer, D. W. (2023). Ordered Network Analysis. In C. Damșa & A. Barany (Eds.), *Advances in Quantitative Ethnography* (Vol. 1785, pp. 101–116). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-31726-2_8

Tausczik, Y. R., Kittur, A., & Kraut, R. E. (2014). Collaborative problem solving: A study of MathOverflow. *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, 355–367. https://doi.org/10.1145/2531602.2531690

Thomas, A. P., Roger, D., & Bull, P. (1983). A sequential analysis of informal dyadic conversation using Markov chains. *British Journal of Social Psychology*, *22*(3), 177–188. https://doi.org/10.1111/j.2044-8309.1983.tb00582.x

Zhang, S., Earle-Randell, T. V., Shen, Q., Botelho, A. F., Israel, M., Boyer, K. E., ... & WIEBE, E. (2024, November). Predicting and Analyzing Students' Higher-Order Questions in Collaborative Problem-Solving. In *International Conference on Computers in Education*. APSCE.

Acknowledgements

We would like to thank the National Science Foundation (#2331379), Institute of Education Sciences (#R305B230007), Gates Foundation (#078981), Support from the Learning Engineering Tools Competition, and other anonymous philanthropy.

International Society of the Learning Sciences

Designing for Collaboration: An Engineering Notebook to Scaffold Collaborative Problem-Solving in Middle School Robotics

Toni Earle-Randell, University of Florida, tearlerandell@ufl.edu
Chris Palaguachi, University of Illinois Urbana-Champaign, cwp5@illinois.edu
Joanne Barrett, University of Florida, jrbarrett@coe.ufl.edu
Jennie Lee, University of Illinois Urbana-Champaign, jl294@illinois.edu
Jina Kang, University of Illinois Urbana-Champaign, jinakang@illinois.edu
Maya Israel, University of Florida, misrael@coe.ufl.edu
Emmanuel Dorley, University of Florida, edorley@ufl.edu
Kristy Elizabeth Boyer, University of Florida, keboyer@ufl.edu
H. Chad Lane, University of Illinois Urbana-Champaign, hclane@illinous.edu

Abstract: Engineering notebooks are pervasive in K-12 robotics as a record of work throughout the competitive season. While these team notebooks are designed to document the engineering design process and students' learning, they have the potential to foster effective collaboration, with students working together rather than in parallel. Capitalizing on this potential, we introduce a new type of robotics (VEX IQ) engineering notebook designed to facilitate collaborative learning for middle schoolers. By integrating pair programming, the Engineering Design Process, and structured prompts for team and code reflection, our notebook serves as an intervention that transforms documentation practice, scaffolds collaborative problem-solving, and makes collaboration a central tenet of the engineering experience.

Introduction

Engineering notebooks serve as a record of work completed by teams of engineers who use documentation to support ongoing development. These notebooks help to scaffold the growth of engineering skills such as the Engineering Design Process (EDP) (Han et al., 2023), increase student engagement (Hertel et al., 2017), encourage reflective thinking (Rodgers, 2002), and facilitate problem-solving during project-based learning experiences (Berland et al., 2012; Mein & Convertino, 2022). However, in a K-12 competitive robotics context, these benefits are often unrealized. While engineering notebooks are critical, providing opportunities to win awards and serving as a guideline for the EDP, many teams struggle to use them as true collaborative tools (Sudano & Avvento, 2024). Common challenges that limit their use as collaborative tools include students working in parallel rather than together, one team member dominating while others disengage, and documentation becoming a retrospective chore rather than an active problem-solving tool. Such collaborative breakdown leads to fragmented understanding, avoidable mistakes, and ineffective communication between team members (Le et al., 2018). Our work aims to address this gap by reimagining the engineering notebook as an active intervention that embeds collaborative practices directly into the documentation of the EDP. Rather than assuming collaboration will occur naturally, we design explicit structures and activities (Weinberger, 2011) that make teamwork unavoidable and productive.

VEX Full Volume challenge

Our VEX engineering notebook was designed as a companion to the VEX IQ "Full Volume" virtual challenge. In this challenge, students work through pair programming activities following the structure provided by the notebook. Students are paired together and asked to program a robot in VEX VR to pick up blocks of different sizes and drop them in various goals, with additional points for strategic decisions. As part of this challenge, students complete ten distinct tasks of increasing complexity. These tasks range from scoring a single block in a goal post (see Figures 1a & 1b) to maximizing points scored in one minute. After these tasks, students are instructed to document their process and reflect on their learnings through the engineering notebook.

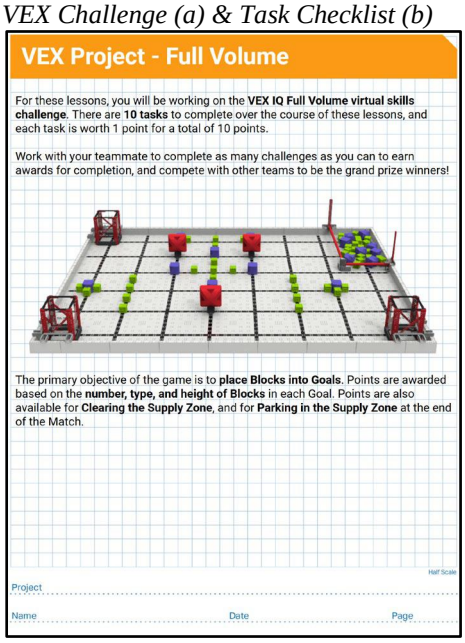
As a metacognitive strategy to scaffold student learning and prevent documentation from becoming a chore rather than a meaningful learning experience, we position documentation as a central component of each task. This is done by grounding our work in engagement and motivation frameworks, particularly self-determination theory (Barata et al., 2013; Ryan & Deci, 2000). We employ gamification strategies by providing bonus points for completing documentation tasks involved in collaborative coding and team reflection activities. Additionally, we leverage the task checklist page to help both the students and instructors monitor progress and completion of tasks (see Figure 1b). Embedding these elements within this engineering notebook reframes

 © ISLS

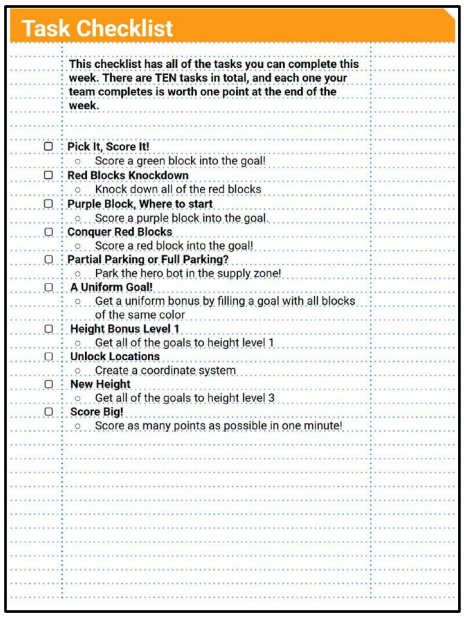
documentation as a goal-directed, collaborative behavior. These activities are intended to support students in tackling increasingly complex tasks by encouraging reflection on prior successes and challenges while fostering key motivational skills such as self-monitoring and self-evaluation in both their coding and group work (Kanfer, 1990).

Figure 1

VEX Challenge (a) & Task Checklist (b)



(a)



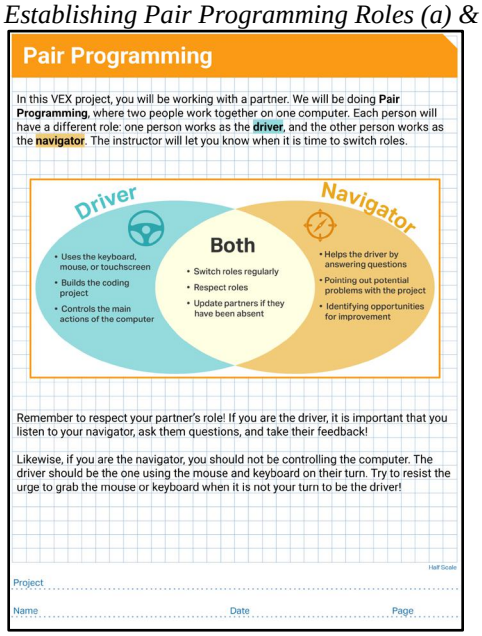
(b)

A notebook designed for collaboration

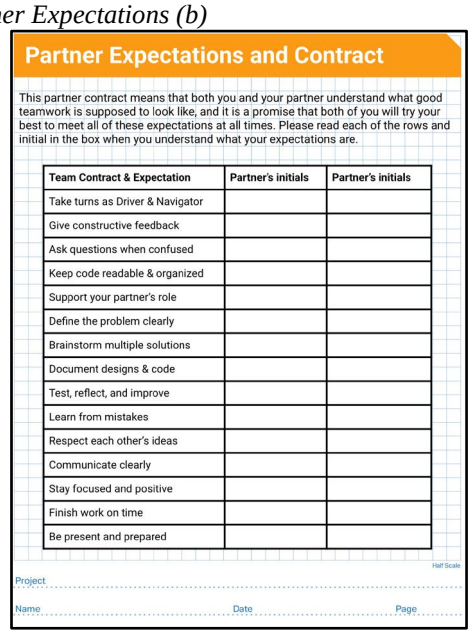
Common collaboration issues that occur in VEX teams are that students work in parallel rather than together, and engagement is often not shared equally between team members. To address these challenges and ensure collaboration throughout the design process, we designed the notebook around three interconnected principles that help to foster genuine teamwork:

- **Structuring the Engineering Design Process Collaboratively -** Our notebook explicitly introduces the EDP principles (Define, Develop, and Test) as an iterative cycle that teams perform together. Each task within the notebook follows this structure, requiring students to collectively define the task goal, develop code as a pair, and jointly test and evaluate their solution. This approach reflects the collaborative nature of contemporary engineering practice, where engineers rarely design in isolation (Borsato & Peruzzini, 2015), while preventing the common pattern where students divide and conquer rather than work together.
- **Establishing Explicit Collaborative Roles -** To prevent the common issue where one student dominates the computer while another disengages, our notebook introduces the Pair Programming methodology that assigns students to one of two roles: the Driver, who controls the computer, and the Navigator, who guides and suggests edits. Students are required to switch roles frequently throughout each task (see figure 2a). By defining these roles explicitly, the notebook provides a clear framework for interaction, encouraging shared responsibility, active listening, and ensuring both students remain engaged in the problem-solving process (Denner et al., 2021).
- **Creating a Shared Team Contract -** Before beginning the project, students sign a "Partner Expectations and Contract" (see Figure 2b), which requires them to agree on principles of good teamwork, such as giving constructive feedback, supporting each other, and respecting each other's ideas. This practice establishes positive group norms from the outset and creates a shared understanding of what successful collaboration entails (Brannen et al., 2021). This contract serves as a reference point throughout the project that students and teachers use during team reflections after each activity.

Figure 2
Establishing Pair Programming Roles (a) & Partner Expectations (b)



(a)



(b)

These structural features create the foundation for collaboration, but the notebook's primary mechanism for facilitating collaborative problem-solving lies in its recurring reflection activities, which we describe next.

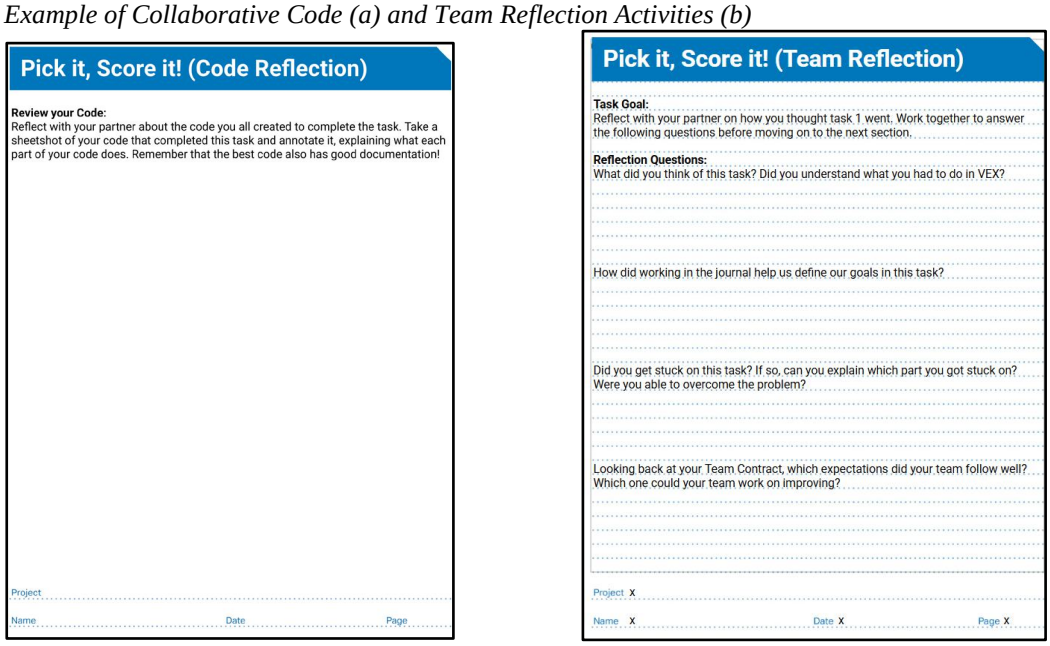
Activities for collaborative problem-solving (CPS)

The notebook's core contribution is its set of recurring activities designed to elicit and scaffold collaborative discussion at key moments in the problem-solving process. These activities align with the "Test" phase of the EDP, ensuring that evaluation and iteration happen collaboratively rather than individually (Sankaranarayanan et al., 2021). Each of the ten tasks is followed by two distinct reflection sections that transform documentation from a passive record into an active site for joint sense-making (Zhang et al., 2024).

Code Reflection: After completing a programming task, students must screenshot their code and collaboratively annotate it to explain what each part does (see Figure 3a for description). This activity requires students to move beyond simply producing functional code. Moreover, the act of generating explanation (self-explanation) is a known strategy shown to support deeper learning in several domains (Rittle-Johnson & Loehr, 2017; Wylie & Chi, 2014), especially in computer programming (Fabic et al., 2020). In addition to a deeper understanding of the content, reflecting on code generated through pair programming can help students come to a deeper shared understanding of the content through discussion and documentation. For example, rather than one student saying "this loop works," both partners must articulate why the loop is needed, how it functions, and what would happen if it were changed. This dedicated space for collaborative reflection on the code transforms documentation from a solitary task into a shared sense-making activity, forcing partners to articulate their logic and build a common understanding of their solution (Sankaranarayanan et al., 2021). It also provides a tangible record of the team's technical solution and the reasoning behind it, which becomes critical for iteration in later tasks when students need to build on previous code.

Team Reflection: Following the code reflection, students encounter a team reflection page with structured prompts about their collaborative process and their experience with the activity (see Figure 3b). These questions include "What was one thing you did well as a team?", "How did working in the journal help your team define your goals in this task?" and "What will you do differently next time to work better together?" These prompts encourage metacognitive discussion about the collaborative process itself, helping students identify both strengths and weaknesses in their communication and problem-solving strategies (Schürmann et al., 2025). By consistently and explicitly prompting students to discuss their teamwork, the notebook elevates the importance of the collaborative process to be on par with the technical solution. Meanwhile, it also provides a structured moment for students to voice concerns, celebrate successes, and strategize ways to improve their partnership in the next task, which develops crucial socio-emotional and communication skills that are central to engineering success (Dahm et al., 2009).

Figure 3
Example of Collaborative Code (a) and Team Reflection Activities (b)



(a) (b)

Together, these two reflection activities create a rhythm of technical work followed by collaborative sense-making, making reflection and iteration an integral part of the engineering process rather than an afterthought.

Reflection, discussion, and next steps

Before developing the current version of the engineering notebook, our research team conducted a series of informal 30-minute interviews with middle school computer science teachers (n=3) and a VEX competition engineering notebook reviewer (n=1). These conversations helped us understand key differences between how engineering notebooks are used for *competition* versus *learning*. Specifically, we found that *competition notebooks* tend to include less scripting and fewer prompts, whereas *teacher-adapted notebooks* for classroom use are typically more structured and guided.

As traditional VEX engineering notebooks are relatively unstructured, for the first version of our notebook we adapted existing VEX materials to add a level of structure. We conducted a usability study at a computer science summer camp, which helped us arrive at a balance between open-ended exploration and structured guidance for the VEX challenge context. Insights from this study indicated that our previous iteration of the notebook did not adequately support productive collaboration or meaningful learning experiences. Instructors reported that there was insufficient prompting to sustain engagement and teamwork. Moreover, students tended to disregard documentation entirely unless it was directly connected to their active task or tied to an explicit incentive. We note that these findings emerged in an informal summer camp setting where students voluntarily participated; the balance of structure and guidance may need adjustment for formal classroom contexts where participation is mandatory and assessment stakes differ.

These insights directly informed the redesign of our engineering notebook present in this study. The new version serves as a design contribution that operationalizes theories of collaborative learning by embedding them into the curriculum itself. Rather than treating the notebook as a passive record, this version positions it as an active scaffold for fostering collaborative problem-solving. It makes collaboration unavoidable by assigning structured team roles, incorporating explicit reflection moments, and linking joint documentation to performance points in the VEX challenge.

As this is a work in progress, we aim to continue collaborating with teachers, students, and competition experts to refine the balance between structure and autonomy in future iterations. Our next phase involves testing this redesigned notebook in middle school computer science classrooms and afterschool programs, documenting challenges, and iterating based on user feedback.

 © ISLS

References

Atman, C.J., Adams, R.S., Cardella, M.E., Turns, J., Mosborg, S., & Saleem, J.J. (2007). Engineering Design Processes: A Comparison of Students and Expert Practitioners. *Journal of Engineering Education, 96*.

Barata, G., Gama, S.P., Jorge, J.A., & Gonçalves, D. (2013). Improving participation and learning with gamification. *Proceedings of the First International Conference on Gameful Design, Research, and Applications*.

Berland, L. K., McKenna, W., & Peacock, S. B. (2012). Understanding Students' Perceptions on the Utility of Engineering Notebooks. *Advances in Engineering Education*, 3(2), n2.

Borsato, M., & Peruzzini, M. (2015). Collaborative engineering. In *Concurrent engineering in the 21st century: Foundations, developments and challenges* (pp. 165-196).

Brannen, S. F., Beauchamp, D., Cartwright, N. M., Liddle, D. M., Tishinsky, J. M., Newton, G., & Monk, J. M. (2021). Effectiveness of group work contracts to facilitate collaborative group learning and reduce anxiety in traditional face-to-face lecture and online distance education course formats. *International Journal for the Scholarship of Teaching and Learning*, 15(2), 5.

Dahm, K., Newell, J., & Newell, H. (2009). The impact of structured writing and developing awareness of learning preferences on the performance and attitudes of engineering teams. *Advances in Engineering Education*

Denner, J., Green, E., & Campe, S. (2021). Learning to program in middle school: How pair programming helps and hinders intrepid exploration. *Journal of the Learning Sciences*, 30(4-5), 611-645.

Fabic, G.V.F., Mitrovic, A. & Neshatian, K. (2020). Evaluation of Parsons Problems with Menu-Based Self-Explanation Prompts in a Mobile Python Tutor. Int J Artif Intell Educ 29, 507–535

Han, J., Park, H. K., & Kelley, T. R. (2023). Engineer's Notebook as a Cognitive Device: Developing a Real-Time Collaborative Engineer's Notebook iOS Application. *Technology and Engineering Teacher*, 82(5).

Hertel, J. D., Cunningham, C. M., & Kelly, G. J. (2017). The roles of engineering notebooks in shaping elementary engineering student discourse and practice. *International Journal of Science Education*, 39(9).

Kanfer, R. (1990). Motivation and individual differences in learning: An integration of developmental, differential and cognitive perspectives. *Learning and Individual Differences, 2*(2), 221–239.

Le, H., Janssen, J., & Wubbels, T. (2018). Collaborative learning practices: teacher and student perceived obstacles to effective student collaboration. *Cambridge Journal of Education*, 48(1), 103-122.

Mein, E., & Convertino, C. (2022). The Uses of Engineering Notebooks Among Pre-Engineering Students at a Hispanic-Serving Institution. *Journal of Adolescent & Adult Literacy*, 66(2), 70-79.

Rittle-Johnson, B., & Loehr, A. M. (2017). Eliciting explanations: Constraints on when self-explanation aids learning. *Psychonomic Bulletin & Review, 24*(5), 1501–1510.

Rodgers, C. R. (2002). Defining reflection: Another look at John Dewey and reflective thinking. *Teachers College Record*, 104(4), 842-866.

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary educational psychology*, *25*(1), 54-67.

Sankaranarayanan, S., Ma, L., Kandimalla, S. R., Markevych, I., Nguyen, H., Murray, R. C., Bogart, C. Hilton, M., Sakr, M., & Rosé, C. P. (2022). Collaborative reflection “in the flow” of programming: designing effective collaborative learning activities in advanced computer science contexts. *In Proceedings of the 15th CSCL 2022*, pp. 67-74. International Society of the Learning Sciences.

Sudano, E., & Avvento, G. (2024). Improving Competition Performance To Middle School Robotic Teams By Teaching Product Development Skills. *Inted2024 Proceedings*, 626-636.

Schürmann, V., Bodemer, D., & Marquardt, N. (2025). Exploring the use of regular reflections in student collaboration: a case study in higher education. *In Frontiers in Education* (Vol. 10, p. 1526487).

Weinberger, A. (2011). Principles of transactive computer-supported collaboration scripts. *Nordic Journal of Digital Literacy*, *6*(3), 189-202.

Wylie, R., & Chi, M. T. (2014). The self-explanation principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (Vol. 2, pp. 413–432). Cambridge University Press.

Zhang, Z., Bekker, T., Skovbjerg, H., & Markopoulos, P. (2024). Supporting and understanding students’ collaborative reflection-in-action during design-based learning. *International Journal of Technology and Design Education.*

Acknowledgments

This material is based upon work supported by the National Science Foundation and the Institute of Education Sciences under Grant DRL2229612. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation or the U.S. Department of Education.

International Society of the Learning Sciences

"It Wasn't So Outside of the Box, But It Wasn't Something I Would Have Thought About": AI as a Creative Collaborator in Student Problem Solving

Ibrahim Oluwajoba Adisa, Stanford University, jobaa@stanford.edu
Victoria Delaney, San Diego State University, vldelaney@sdsu.edu

Abstract: This study examines how students experience collaboration with generative AI during creative problem-solving. Using post-interview data, we analyze students' perceptions of creativity and agency across three human–AI writing modes. Findings show that students treated AI as a responsive collaborator that expanded ideas and supported refinement, but was most valued for structuring and improving efficiency. The richest collaborations occurred when students remained active in both generating and refining ideas. These results highlight how generative AI differentially supports divergent and convergent thinking and suggest that balanced, iterative interaction is key to preserving agency in human–AI co-creation.

Introduction

Creativity is widely understood as a universal human capacity involving the ability to imagine possibilities, connect disparate ideas, and produce work that is both novel and meaningful (Runco & Jaeger, 2012). Far from being limited to the arts, creativity plays a central role across the sciences, humanities, and professional domains (Brown & Kātz, 2009; Kaufman & Beghetto, 2009). Given its importance, researchers have long sought to understand the cognitive mechanisms that support creative thinking and how emerging technologies might scaffold the creative process (Lubart, 2005; Sawyer & Henriksen, 2024). One promising area of inquiry is the use of generative AI as a collaborative partner in creative writing.

Creative writing involves imaginative thinking, emotional expression, and the deliberate crafting of language that draws on imagery, rhythm, and metaphor to give form to abstract ideas (Hanauer, 2012). Writing a poem, for example, typically unfolds through iterative phases of idea generation, exploration, revision, and refinement, processes that align with established models of creative problem solving (Guilford, 1967; Treffinger & Isaksen, 2005). Across these phases, writers move fluidly between divergent thinking, generating multiple possibilities, and convergent thinking, selecting, evaluating, and organizing ideas into a coherent composition. Because creative writing integrates affective, cognitive, and aesthetic dimensions, it provides a rich context for examining how AI may mediate or extend human creativity. Unlike earlier tools that primarily automated routine tasks, generative AI systems can engage interactively by offering alternative perspectives and linguistic variations that can stimulate new lines of thought, while the writer retains control over which suggestions to adopt, revise, or reject.

These capabilities, however, raise important questions about AI's role in creative work and how learners negotiate agency within human–AI partnerships. Recent studies suggest that AI can enhance originality and fluency by expanding idea spaces and supporting creative risk-taking (Davis et al., 2024). Others caution that heavy reliance on AI may reduce cognitive engagement, lead to formulaic outputs, or diminish individuals' sense of ownership (Dhillon et al., 2024; Frich et al., 2024). Examining students' iterative collaboration with generative AI during creative writing tasks provides a way to understand how creativity emerges through this human–AI partnership, including how ideas are generated, taken up, and transformed across different configurations of human and AI contribution. In this study, we investigate how students experience collaboration with AI in creative poem-writing tasks that vary in their balance of human and AI contributions. Specifically, we ask: *How do students experience and perceive collaboration with generative AI during creative problem-solving tasks?*

Creative problem solving and human-AI co-creation

Creative problem solving involves applying creative thinking within the purposeful act of addressing a challenge or creating a product (Treffinger et al., 2023; Treffinger & Isaksen, 2005). Central to creativity is the dynamic interplay between divergent and convergent thinking (Guilford, 1967; Cropley, 2006). Divergent thinking supports exploration and the generation of multiple ideas, while convergent thinking enables the evaluation and integration of those ideas into coherent outcomes. Creativity flourishes when individuals can fluidly move between these two modes, expanding and narrowing ideas through iterative cycles (Runco & Acar, 2012). Poetry writing exemplifies this process, requiring cycles of idea generation, revision, and reorganization to communicate

 © ISLS

meaning and evoke emotion, while also supporting metaphorical reasoning, emotional expression, and perspective taking (Faulkner, 2023; Hanauer, 2012; Januchowski-Hartley et al., 2018).

Because creativity rarely occurs in isolation, researchers have increasingly highlighted its social and collaborative dimensions (Paulus & Nijstad, 2019; Sawyer & Henriksen, 2024). Collaboration allows individuals to access diverse perspectives, critique ideas, and co-construct solutions (Paulus & Nijstad, 2019; Sawyer & Henriksen, 2024). In human–AI co-creation, these collaborative processes are reconfigured, as AI systems contribute suggestions, analogies, and structural possibilities, while humans interpret, evaluate, and contextualize them (Davis et al., 2024). Prior work suggests that the level of AI support plays a critical role in shaping both creative outcomes and learners' sense of ownership. While more substantial AI contributions can improve quality and productivity (Dhillon et al., 2024; Huang, 2025), it is unclear whether students who partake in creative and personal endeavors (such as writing poetry) consider the productivity gains of AI to offer authentic improvements to their creative processes. From a learning perspective, collaboration between learners and AI raises questions of epistemic agency, including who generates ideas, evaluates contributions, and determines what counts as a satisfactory outcome (Scardamalia & Bereiter, 2020). Additionally, students' perceptions of AI, including its capabilities, trustworthiness, and perceived "social" role, further shape how they engage with it, and influence whether they critically probe, adapt, or defer to its suggestions (Rezwana & Ford, 2025). As such, examining how students collaborate with AI during poetry writing provides a focused lens for understanding how creative processes, agency, and co-creation unfold in human–AI interaction.

Methods

This exploratory qualitative study (Creswell & Poth, 2018) examined how students experience and perceive collaboration with generative AI, specifically ChatGPT, during creative problem-solving. The study was conducted virtually over Zoom in summer 2025 across three consecutive days, with one-hour sessions each day. Students engaged in poem-writing tasks under three modes of human–AI collaboration. The tasks were designed to be comparable in difficulty while varying in social context to support creativity without privileging any session. Seven students (all girls), aged 11–16, participated (four middle school, three high school). This all-female composition reflects the available participant pool for this pilot and is not intended to represent broader gendered patterns. All participants completed the three collaboration modes, engaging with AI as a human–AI dyad. All reported prior experience with generative AI and familiarity with basic prompting. Sessions were facilitated by the authors, who introduced prompting strategies and guided students through each activity. Parental consent and participant assent were obtained in accordance with university ethical guidelines.

The first session introduced generative AI and prompt engineering, including strategies such as clarity, context, and iterative refinement. Students also completed a brief "prompt detective" activity to highlight how prompts shape AI outputs. Students then completed three creative tasks reflecting different collaboration modes, ranging from more student-directed to more AI-mediated engagement, each emphasizing different phases of divergent and convergent thinking. In *AI-supported expansion*, students independently brainstormed and drafted a poem on teen mental health, then used ChatGPT to generate additional ideas before independently synthesizing both sources into a final piece. This mode foregrounded divergent thinking by expanding the range of ideas available to students while still requiring them to evaluate and integrate these ideas into a coherent product. In *AI-supported synthesis*, students independently brainstormed ideas for a poem on belonging or pressure, then used ChatGPT to organize their ideas into a structured plan before composing their final poem. Here, AI primarily supported convergent thinking by helping students evaluate, organize, and refine their ideas into a cohesive structure. In *AI-initiated drafting*, students prompted ChatGPT to generate a poem about wildfire safety, iterated on prompts to refine outputs, and optionally revised the final product to personalize it. This mode shifted much of the initial idea generation to the AI, positioning students to engage more critically in convergent processes such as evaluating, selecting, and modifying AI-generated content. These conditions allowed us to examine how varying levels of AI involvement shaped students' engagement in divergent and convergent thinking, as well as their perceptions of collaboration, authorship, and creative ownership.

Data sources and analysis

Data sources included a pre-survey, three screen recordings of sessions, seven chat logs, and two group post-interviews (approximately 102 minutes total). The pre-survey captured prior AI experience. Screen recordings documented students' reflections after each task. Chat logs captured prompts and interaction strategies. Post-interviews, conducted over Zoom in groups of three to four students, focused on students' perceptions of AI's role in their creativity, originality, and efficiency. For this paper, we draw primarily on post-interview data, using other sources to contextualize students' reflections. Interviews were transcribed using Zoom's automated service and manually reviewed for accuracy. We conducted an inductive, two-cycle coding process (Saldaña, 2021)

informed by our focus on collaborative and creative problem solving. In the first cycle, we generated 15 descriptive and value-oriented codes (e.g., AI as partner, preference for personal authorship, appreciation of AI support). In the second cycle, we grouped these into broader categories related to perceptions of AI as collaborator, its role in shaping creativity and originality, and students' sense of authorship and ownership. Our analysis attended to both within-mode patterns and cross-mode comparisons to examine how different levels of AI involvement shaped students' experiences and perceptions.

Findings

Experiencing AI as a responsive and respectful collaborator

Across the three collaborative modes, students interacted with AI as if they were in dialogue with a creative partner that could listen, respond, and adapt to their intentions. They often spoke about AI as having a sense of awareness and sensitivity that shaped the rhythm of collaboration. Several students were surprised by what they perceived as empathy or emotional connections in the AI's responses. For example, Amy (pseudonym) remarked, "It was interesting how it [AI] was able to capture human feelings… it kind of captured the essence of what a human could be going through." Others appreciated AI's ability to honor conversational boundaries and respect instructions. Emma described how working with the AI in the first two modes felt like working with a peer because it could remember and sustain shared understandings across interactions. She elaborated: "One thing that was surprising was that I would tell the AI not to write the actual poem … and I was skeptical if the AI would be able to remember … and surprisingly it did." Students interpreted this consistency as evidence that the AI was participating in a mutual process, holding context and maintaining respect for their creative space. Students also appreciated how the AI remained responsive even when their prompts were unclear. Eva explained, "Even if my question was kind of confusing, it still answered my question the way I wanted it to." For her, this quality mirrored a supportive peer who strives to understand or anticipate her thinking process. These interactions reflect students' positioning of AI as a collaborative partner in co-creation, where meaning emerges through iterative dialogue rather than one-sided generation.

AI appeared most helpful for refinement, structure, and efficiency

When describing how AI shaped their creative process, students emphasized that it contributed new perspectives and stylistic suggestions that expanded their ideas. During the AI-supported expansion mode, they highlighted AI's capacity to provide "solutions I did not think about," generate evocative metaphors, and refine the tone and rhythm of their poems. Tracy shared that while she generated most of her own ideas in the first two collaboration modes, AI "made my writing a bit clearer, but without losing my original vision." Eva echoed this, stating that AI made her work "more original because it gave me… unique ideas that I didn't really think of before." These comments suggest that students viewed AI as a creative partner that provided access to ideas and expressions beyond their usual imaginative range. In this way, AI-supported expansion primarily contributed to divergent thinking by broadening the range of ideas and perspectives available to students. Yet not all students shared this sense of the AI providing unique contributions. Some remained cautious about the originality of AI-generated ideas. For instance, Amy noted that while the AI's ideas helped her see the "big picture" and organize her thinking, they were not "completely out of the box." Across participants, AI was most valued for widening the range of creative possibilities and helping them articulate ideas they already held, rather than for inventing entirely novel concepts.

Although students appreciated AI's ability to extend the fluency and flexibility of their ideas, many noted that their most productive exchanges with AI occurred during convergent phases, when both partners worked to synthesize and bring structure and meaning to their shared ideas. Many described brainstorming as "pretty easy," but recognized AI's contribution in "putting it all together" and helping structure their poems more cohesively in the second mode. Amy explained, "It was pretty easy for me to brainstorm and come up with bullet points of ideas. But I think where AI really became helpful was like putting it all together … It helped me structure it better." Emma similarly noted that AI's primary role was "refining" her ideas, while Eva appreciated the constructive feedback and editing support it provided. Students described their final outputs as "more polished," "proper," and "professional." Importantly, students who reported lower confidence in poetry writing found AI especially helpful during brainstorming and synthesis. Tracy admitted, "I was having problems finding rhyming words… [because] I'm not that great at writing poems [but] AI helped me." Efficiency also emerged as a recurring benefit. Across tasks, students noted that writing "took way less time with AI." These patterns suggest that students perceived AI not only as a source of ideas but as a tool that more strongly supported convergent processes of refinement, coherence, and efficiency within creative problem solving.

Negotiating control, quality, and collaboration

Students' experiences also highlighted the negotiation of epistemic agency within human–AI co-creation, particularly in how control over idea generation and refinement shifted across modes. While students valued AI's feedback and efficiency, they also encountered challenges in maintaining creative control. Some reported moments when AI's suggestions diverged from their intentions. Tracy described having to "redirect it and give it something different" when AI's initial feedback did not align with her vision. Eva also noted that AI-generated poems were "shorter" and lacked "small little details" compared to those she co-created more actively. She explained: "My task three [poem], which was mostly the AI, was a shorter poem … in task two and one, it was kind of a longer poem just because that was more like my ideas." For this student, efficiency came at a creative cost, with the poem being cleaner but lacking emotional depth. Additionally, when students struggled to articulate their thoughts in prompts, they expressed frustration, with Nora noting, "I wish it could just like read my mind because sometimes I didn't know how to put the things I was thinking into words." Students differed in which collaboration mode they experienced as most balanced, and their sense of collaboration was shaped more by the quality of interaction than by the amount of AI contribution. Amy felt AI was most collaborative in the second mode when she "spent more time explaining [her] thinking and giving feedback," describing that process as when the poem "felt the most like what I was trying to write." For her, the alternation between generating and refining ideas created a "50-50" partnership. Eva and Emma, on the other hand, felt more collaborative in the third mode because they "prompted it a lot more" and saw the AI contributing more ideas and revisions. Conversely, Tracy felt most partnered with AI in the first mode, when she and the AI "fed off from each other." When reflecting on their poems, students consistently linked creative quality to their perceived share of effort and emotional investment. Eva described her first poem as her best because it had "a lot of deep meaning," while Emma preferred her second poem for showing "a lot of my creativity and imagination." Despite appreciating the clarity of AI-generated work in the third mode, most agreed that their earlier, more self-driven poems in the first two modes felt "more creative" and "more complex." These findings suggest that students' sense of collaboration was shaped less by how much the AI contributed and more by how interaction supported their agency in both generating and shaping ideas.

Discussion and conclusion

This study examined how students experienced collaboration with generative AI during creative problem-solving, focusing on how different collaboration modes shaped their engagement in divergent and convergent thinking. Across modes, students treated AI as a responsive partner that could extend ideas and support refinement, while also revealing tensions in how AI contributed to, and sometimes constrained, the creative process. A key contribution of this work is showing how generative AI differentially supports phases of creative problem solving. Consistent with CPS frameworks (Guilford, 1967), students' independent and AI-supported idea generation reflected divergent thinking, while their use of AI for structuring, refining, and editing aligned with convergent thinking. Although AI-supported expansion broadened idea spaces, students consistently found AI most valuable during convergent phases, where it helped organize, clarify, and finalize their work. This suggests that, in practice, AI functioned less as a source of entirely novel ideas and more as a tool for shaping, refining, and making ideas usable. This tension is reflected in students' own accounts of AI as "not completely out of the box," yet still generative of ideas they "would not have thought about."

The three collaboration modes further revealed how shifts in AI involvement redistributed students' roles. In AI-supported expansion and synthesis, students retained responsibility for generating and shaping ideas, with AI extending or organizing their thinking. In contrast, in AI-initiated drafting, students were positioned more as evaluators and editors. While this increased efficiency and coherence, outputs were often perceived as less original, detailed, and personally meaningful. This helps explain why more balanced modes were experienced as more creative and collaborative. Creativity, in this context, depended less on the amount of AI contribution and more on students' active engagement in both generating and refining ideas. These findings complicate claims of AI as an equal creative partner. While students described AI in relational terms, this reflects an interactional experience rather than an equivalent distribution of agency. Instead, AI functioned as a form of adaptive scaffolding, enabling students to move more fluidly between divergent and convergent processes by offering just-in-time suggestions, alternative framings, and structural support. From a design perspective, these findings highlight the importance of structuring human–AI interaction to preserve student agency while leveraging AI support. Tasks that require students to generate initial ideas, engage in iterative exchanges, and evaluate AI contributions appear to better support creative problem solving than those that position AI as the primary generator of content. This exploratory study is limited by a small, all-female sample, the use of a single AI tool, and non-controlled collaboration modes, but it offers early insight into how different human–AI configurations shape students' experiences of creativity, agency, and authorship.

References

Brown, T., & Kātz, B. (2009). *Change by design: How design thinking transforms organizations and inspires innovation* (1st ed.). Harper Business.

Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches.* Sage.

Cropley, A. (2006). In praise of convergent thinking. *Creativity Research Journal, 18*(3), 391–404.

Davis, N., Deshpande, M., Rezwana, J., & Magerko, B. (2024). The Five Pillars of Enaction as a Theoretical Framework for Co-Creative Artificial Intelligence.

Dhillon, P. S., Molaei, S., Li, J., Golub, M., Zheng, S., & Robert, L. P. (2024, May). Shaping human-AI collaboration: Varied scaffolding levels in co-writing with language models. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (pp. 1-18).

Faulkner, S. L. (2023). Setting the agenda: Poetic inquiry as critical family communication methodology. *Journal of Family Communication, 23*(1), 75-87.

Frich, J., Grønbæk, J. E. S., Borowski, M., & Dalsgaard, P. (2024, November). Exploring the impact of AI features on collaborative creativity. In *Proceedings of the 36th Australasian Conference on Human-Computer Interaction* (pp. 480-489).

Guilford, J. P. (1967). *The nature of human intelligence*. McGraw–Hill.

Hanauer, D. I. (2012). Meaningful literacy: Writing poetry in the language classroom. *Language Teaching*, *45*(1), 105–115. https://doi.org/10.1017/S0261444810000522

Huang, H. C. B. (2025). Unlocking creativity with artificial intelligence (AI): Field and experimental evidence on the Goldilocks (curvilinear) effect of human–AI collaboration. *Journal of Experimental Psychology: General*.

Januchowski-Hartley, S. R., Sopinka, N., Merkle, B. G., Lux, C., Zivian, A., Goff, P., & Oester, S. (2018). Poetry as a creative practice to enhance engagement and learning in conservation science. *BioScience, 68*(11), 905-911. https://doi.org/10.1093/biosci/biy105

Kaufman, J. C., & Beghetto, R. A. (2009). Beyond big and little: The four c model of creativity. *Review of General Psychology, 13*(1), 1–12. https://doi.org/10.1037/a0013688

Lubart, T. (2005). How can computers be partners in the creative process: Classification and commentary on the special issue. *International Journal of Human-Computer Studies, 63*(4–5), 365–369.

Paulus, P. B., & Nijstad, B. A. (Eds.). (2019). *The Oxford handbook of group creativity and innovation*. Oxford University Press.

Rezwana, J., & Ford, C. (2025, June). Human-Centered AI Communication in Co-Creativity: An Initial Framework and Insights. In Proceedings of the 2025 Conference on Creativity and Cognition (pp. 651-665).

Runco, M. A., & Acar, S. (2012). Divergent thinking as an indicator of creative potential. *Creativity Research Journal, 24*(1), 66–75. https://doi.org/10.1080/10400419.2012.652929

Runco, M. A., & Jaeger, G. J. (2012). The standard definition of creativity. *Creativity Research Journal, 24*(1), 92–96. https://doi.org/10.1080/10400419.2012.650092

Saldaña, J. (2021). The coding manual for qualitative researchers (4th ed.). Sage Publications.

Sawyer, R. K., & Henriksen, D. (2024). *Explaining creativity: The science of human innovation*. Oxford university press.

Scardamalia, M., & Bereiter, C. (2020). Will knowledge building remain uniquely human? *QWERTY-Interdisciplinary Journal of Technology, Culture and Education, 15*(2), 12-26..

Treffinger, D. J., & Isaksen, S. G. (2005). Creative problem solving: The history, development, and implications for gifted education and talent development. *Gifted Child Quarterly, 49*(4), 342–353. https://doi.org/10.1177/001698620504900407

Treffinger, D. J., Isaksen, S. G., & Stead-Dorval, K. B. (2023). *Creative problem solving: An introduction.* Routledge.

Acknowledgments

We gratefully acknowledge Self-eSTEM for their support of this research, including facilitating access to participants and enabling the implementation of this study.

International Society of the Learning Sciences

Are We There Yet? The Roles and Development of GenAI-Based Tools Supporting Collaborative Learning: A Scoping Review

Maya Persico, University of Hamburg, Germany, maya.persico@uni-hamburg.de
Daeun Hong, Indiana University Bloomington, United States, dh37@iu.edu
Jerika Miller, Indiana University Bloomington, United States, jm198@iu.edu
Cindy E. Hmelo-Silver, Indiana University Bloomington, United States, chmelosi@iu.edu
Freydis Vogel, University of Hamburg, Germany, freydis.vogel@uni-hamburg.de

Abstract: Tools based on generative AI (genAI), interacting with learners in a human-like manner, such as conversational agents and ChatGPT, offer potential to develop adaptive support for collaborative learning. Following PRISMA guidelines, after screening results of searching three databases, our scoping review identified 239 relevant records reporting about genAI tools in CSCL including 138 empirical articles this paper focuses on. They were analyzed in depth regarding genAI tool's maturity level of implementation in collaborative learning, and the role genAI takes within. Results show studies at different levels of maturity in implementation of the genAI tool ranging from designing the tool or investigating its usability to studies focusing on human learning outcomes. GenAI takes multiple roles simultaneously, mostly providing epistemic and metacognitive scaffolding to individuals or groups. Our results indicate substantial development in the growing number of empirical articles, while pointing out gaps in specific support type provision, especially for school learners.

Introduction

While collaborative learning is a widely used pedagogical approach, learners do not necessarily engage in productive learning activities autonomously. Thus, there is a long-lasting research tradition developing instructional approaches or scaffolding, often implemented with digital technologies, to facilitate the collaborative learning processes (Wise & Schwarz, 2017). In this vein, researchers in the field of CSCL have been working on the development and investigation of adaptive support for collaborative learning with digital tools (Hmelo-Silver & Jeong, 2021) such as intelligent tutoring systems for peer feedback (Walker et al., 2014). Such adaptive systems can be based on automated discourse analysis, performing analysis and adjusting support in real time (Bae et al., 2019; Rosé et al., 2008). Since digital tools based on generative AI (genAI) and large language models (LLMs) became widely available, new opportunities have emerged to develop adaptive support for beneficial CSCL processes (Cress & Kimmerle, 2023, Yan et al., 2024). By reviewing research on the use of genAI in CSCL, our scoping review provides an overview about existing research and gaps, regarding genAI-based collaborative learning support.

Theoretical background

The developments of integrating genAI into learning environments enable new forms of supporting collaborative learning. For instance, genAI-based tools can take on a range of roles substituting support that had so far been mostly offered by humans, such as tutors or peers, or digital tools based on more straightforward algorithms (Yan et al., 2024). In CSCL learning environments different types of support can be distinguished. Epistemic support provides content information or guides learners to build understanding about a subject (Weinberger et al., 2005). Social support structures and guides learners through social activities (Vogel et al., 2017). Metacognitive support can provide guidance for planning, monitoring, reflection, or conversation on individual and group levels (Järvelä et al., 2023). Support can also be geared toward socio-emotional or motivational aspects emerging in group exchange or it can be used to help teachers for classroom orchestration of CSCL (Bae et al., 2019). For example, argumentation scripts (Noroozi et al., 2013) and group awareness tools (Schnaubert & Bodemer, 2022) combine social and metacognitive support in collaborative learning, opening the field to socially shared regulation (Järvelä et al., 2023). Collaborative learning often consists of alternating individual and collaborative learning phases and may lead to differences to whom the support is provided. Using metacognitive support as an example, tools supporting collaborative learning can target self-regulated learning (SRL), co-regulation (CoRL) or socially shared regulation of learning (SSRL) (Järvelä et al., 2016). What stands out is that genAI may provide these types of timely support both separately and simultaneously, highlighting its potential to support learners and teachers across diverse tasks and contexts. This support can have diverse functions such as a peer, a tutor, or a co-creator, engaging dynamically with learners rather than serving as a static tool (Yan et al., 2024).

 © ISLS

Since the release of the highly capable genAI-based conversational agent ChatGPT 3.5 in November 2022, many publications have appeared regarding the use of genAI in CSCL. Due to its novelty, it is not surprising that the level of development of these tools in research is as diverse as their research objectives are. Researchers started using the new tools with different objectives: testing the tool's abilities, observing learners' use cases and perceptions, or investigating the tools' influence on learning processes. Hence, the maturity level of genAI tool implementation within learning settings ranges on a spectrum from development to pilot to full implementation into CSCL. To provide an overview of what has been published and how genAI has been used, we conducted a scoping review guided by three research questions (RQs). With regard to the whole scope of published papers: (RQ1) Which type of papers have been published about the use of genAI in CSCL since November 2022? With a particular focus on the empirical articles: (RQ2a) What aspects of CSCL are supported by and (RQ2b) who are the support recipients of genAI tools used in the identified papers about using genAI in CSCL? (RQ3) What are the maturity levels of genAI tool's implementation into CSCL?

Methods

Following the PRISMA-ScR guidelines, we conducted several steps to identify studies for inclusion in the scoping review. We searched three databases, Scopus, Web of Science, and ERIC, to find relevant articles. The search was restricted to papers published after the release of ChatGPT in 2022. A consistent set of keywords and phrases, adapted to each database's query input rules, was used in all searches:

> "("learning" OR "knowledge building" OR "knowledge co-construction" OR "knowledge acquisition") AND ("large language model" OR "generative AI" OR "ChatGPT" OR "conversational agent" OR "chatbot") AND ("collaborat*" OR "cooperat*" OR "CSCL")"

To broaden coverage, we included various alternative keywords for "learning" to retrieve papers that might not include the term learning yet still address education and learning. The search yielded 1,132 potential articles published between 2022 and September 5, 2024, when the search was initiated. Removing duplicates reduced the dataset to 856 articles. In total, three rounds of coding were performed.

The first round focused on applying the inclusion and exclusion criteria to determine relevance, first on the basis of abstracts, then based on the entire paper. In the subsequent coding rounds, inclusion and exclusion criteria were further applied based on additional information. The three inclusion criteria comprised whether the articles (1) pertain to human learning or research in an educational context, (2) include genAI or LLMs, and (3) involve human learner collaboration. The five exclusion criteria comprised (1) languages other than English, (2) restricted retrieval, (3) published as a list of papers or as a whole book, (4) published before November 2022, and (5) focusing on interdisciplinary researcher collaboration, or technical systems collaborating with each other. Two coders trained until they reached acceptable inter-rater reliability ($\kappa > 0.75$) for each coding category, using 20% of the sample. Each coder subsequently reviewed half of the residual articles independently, resulting in 239 relevant articles identified for the scoping review. For the second round of coding, we developed and both coders applied a coding scheme to the 239 included articles categorizing the publications into reviews, empirical, conceptual or discussion papers as well as editorials ($\kappa = .662$) (RQ1), solving discrepancies through discussion.

The third round served to answer RQ2 and RQ3, focusing on empirical articles only ($n = 138$). We developed another coding scheme to categorize the empirical papers into 3 dimensions. Two coders double coded the whole sample for each dimension with acceptable inter-rater reliability ($\kappa_{\varnothing} = .771$, $ICC_{\varnothing} = .828$). Discrepancies in coding were resolved through discussion and mutual agreement. The aspects of CSCL support (RQ2a) and support recipient (RQ2b) were operationalized by dimensions 'type of support', 'support recipient' and 'educational level of participants'. More than one type of support could be coded simultaneously. The type of support provided by genAI for collaborative learning consisted of: epistemic/conceptual, social/collaboration, metacognitive/regulation, socio-emotional/motivational, and orchestration support category. The coding rules are presented in Table 1. The support recipients were categorized into individual, group, both, unspecified and other (e.g., for orchestration). For example, in think-pair-share studies we searched in methods and procedure for which parts the learners were allowed to use the tool. If it was provided for individual preparation, we categorized it as individual support; if it was provided support during sharing, we categorized it as group support; if it was provided throughout the learning process for individual brainstorming as well as sharing, we categorized it as 'both'. Beyond that, we coded the educational level of the support recipients distinguishing between university students, other adult learners, school learners, kindergarten and pre-school. The tool's level of maturity in implementation into the learning setting (RQ3) was operationalized with an inductively built hierarchical categorization. Mostly dependent on the reviewed study's research focus (e.g., tool performance focus, usability, learning outcomes), the tools' level of maturity in implementation differed. Research focusing on tool performance was mostly found in

 © ISLS

design process and coded as ‘in-development’, whereas research that focused on usability and usage perceptions of learners that experimented with genAI-based tools was categorized as ‘pilot.’ Studies concerned with human learning outcomes, that integrated AI tools in genuine collaborative learning scenarios, received the highest category with ‘implemented’. Furthermore, descriptive sample details about the used AI’s name were coded.

Table 1
Coding Scheme for Support Type Provided by GenAI in Collaborative Learning

Category	Rule
Epistemic / conceptual (disciplinary)	Provide content-related information, resources, and materials; support students’ understanding and epistemic practice; or giving information about right or wrong.
Social / Collaboration	Support and/or facilitates students’ social interactions including collaboration, communication, participation, and discussions.
Meta-cognitive / Regulation	Support students to monitor and evaluate their learning progress and plan and regulate their learning.
Socio-emotional / Motivation	Support students in remaining engaged and motivated; and foster positive feelings and help manage their emotions during learning.
Orchestration (of learning and environment)	Provides support creating learning environments without directly intervening in students’ learning processes.

Results

Overall sample (n = 239)

Starting with descriptive information about the article sample, most of the relevant articles (n = 239) were published in 2024 (64%). Empirical papers comprise about 58 percent of the entire sample, followed by discussion (20%), conceptual (11%), review articles (10%), and editorials (1%).

Empirical articles (n = 138)

We first present descriptive details about this sample of empirical articles (n = 138). The research methods applied in these studies are quantitative only (32%), qualitative only (22%), using both quantitative and qualitative methods (19%), using explicitly mixed-methods (12%), and design-based-research (15%). ChatGPT and its models were the dominating genAI-based tools used in the studies, comprising about 49 percent.

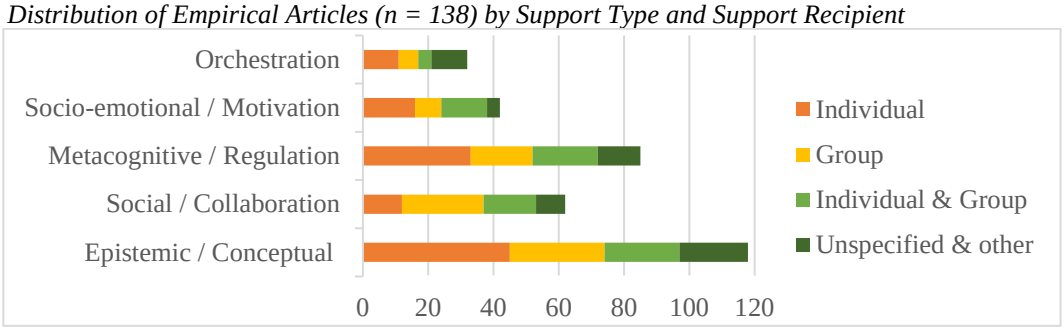
GenAI is being used to support a variety of CSCL aspects. The greatest proportion of support type provided in collaborative learning is epistemic/conceptual, followed by metacognitive support. The genAI-based tools were mostly provided to individuals while learning collaboratively (36%), followed by supporting the whole group (26%), or individuals and groups simultaneously (17%). In most studies the participants were university students (59%) or other adult learners (15%), followed by school learners (9%), and none of the studies reported about participants from kindergarten or pre-school. Around 14 percent included participants of multiple educational levels and 3 percent did not provide information on educational levels of their participants. Figure 1 shows how support recipients are distributed within the support type categories. We note that more than one support type could be provided.

A few examples of the actual implementation may reveal what lies behind the categorization. In one example, support on an individual level is provided by ChatGPT 3.5 as rewriting tool to improve lab report writing: after students finished writing their drafts individually, they had to use the genAI tool for a co-writing process, what increased knowledge of academic writing norms and audience awareness (Kim et al., 2024). Another study on human-AI collaboration patterns found differences in performance, depending on the actual use and (metacognitive) support of genAI-assistance in academic writing (Nguyen et al., 2024). Supporting students on an individual level in sharing their own thoughts (Liu et al., 2015) and scaffolding transactive group exchange (Noroozi et al., 2013), de Araujo and colleagues (2024) similarly combined individual support with group level support in their collaborative conversational agent’s design to assess its impact on student dialogue. One article that was coded for all support types, investigated how to design educational chatbots for collaborative learning. Presenting itself as team member and tutor, the designed Chatbot guides each student group through a learning journey by scaffolding all of our support type categories (Burkhard et al., 2022).

Figure 1

 © ISLS

Distribution of Empirical Articles (n = 138) by Support Type and Support Recipient



With regard to the genAI tool's level of maturity in implementation into CSCL, around 23 percent of the tools used in the empirical studies were still 'in-development', focusing on tool performance and improvement for further development. More articles were in 'pilot' (33%), as they mostly provided a genAI-based tool to participants without focusing on specific learning outcomes, but for testing perceptions of usage. Remarkably, in 44 percent of empirical articles the genAI tool was 'implemented' into the CSCL setting with a focus on human learning outcomes, collaborative learning processes or learning performance improvement. In total, 56 percent of the tools used in the empirical articles have not yet moved beyond pilot implementation, showing the potential for studies to be conducted and published in the future that have genAI-based tools fully implemented.

Discussion and implications

The analysis showed that genAI-based tools have mostly been used for epistemic support in collaborative learning, for instance, to clarify questions, explain concepts, or to generate content (e.g., in form of images, Trizedes et al., 2024). This application does not yet differ far from older CSCL approaches (Vogel et al., 2017). Although some studies still focus quite heavily on one type, all support types were presented within our findings. Our results show numerous studies that take advantage of the new applications, specifically integrating and exploring the diverse and simultaneous use of support by genAI. Although research about adaptive collaboration support has been ongoing for some time now (Bae et al., 2019; Rosé et al., 2008, Walker et al., 2014), only the latest developments have made adaptive scaffolding in real time possible (e.g., Tai & Chen, 2024). The support provision differs in direction to individuals to prepare for collaborative work, to the whole group, or both. Despite the collaborative learning context, AI support is more often directed toward individuals than groups, and more often toward adults than K-12 learners, representing important limitations in current research. The relatively small number of articles reporting metacognitive support at group level indicates a research gap, leaving great opportunities for further research, for example on genAI-based co- and shared-regulation support, to enhance the theoretical and practical understanding of regulation in collaborative learning (Järvelä et al., 2023). Despite the field's recent emergence, the scoping review revealed a large number of empirical articles, suggesting rapid development and progress in research.

Although the number of studies having genAI tools implemented into CSCL environments is already surprisingly large given the short time since ChatGPT 3.5 was released, the number of articles with genAI tools 'in-development' leaves ample opportunity for further research and levels of development beyond the piloting stage in the future as more research is conducted. The results of this scoping review open the door to deeper analyses of the learning outcomes these AI-supported interventions produce, which will be crucial for understanding their real impact in collaborative learning contexts.

References

Bae, H., Glazewski, K., Hmelo-Silver, C., Lester, J., Mott, B. W., & Rowe, J. (2019). Intelligent Cognitive Assistants to Support Orchestration in CSCL. In Lund, K., Niccolai, G. P., Lavoué, E., Hmelo-Silver, C., Gweon, G., & Baker, M. (Eds.), A Wide Lens: Combining Embodied, Enactive, Extended, and Embedded Learning in Collaborative Settings, *13th International Conference on Computer Supported Collaborative Learning (CSCL) 2019*, 2, 947-948, https://doi.dx.org/10.22318/cscl2019.947

*Burkhard, M., Seufert, S., Cetto, M., & Handschuh, S. (2022). Educational Chatbots for Collaborative Learning: Results of a Design Experiment in a Middle School. *19th International Conference on Cognition and Exploratory Learning in Digital Age (CELDA 2022)*, 209–218.

Cress, U., & Kimmerle, J. (2023). Co-constructing knowledge with generative AI tools: Reflections from a CSCL perspective. *International Journal of Computer-Supported Collaborative Learning*, *18*(4), 607–614. https://doi.org/10.1007/s11412-023-09409-w

 © ISLS

*de Araujo, A., Papadopoulos, P. M., McKenney, S., & de Jong, T. (2024). A learning analytics-based collaborative conversational agent to foster productive dialogue in inquiry learning. *Journal of Computer Assisted Learning, 15*. https://doi.org/10.1111/jcal.13007

Hmelo-Silver, C. E., & Jeong, H. (2021). An Overview of CSCL Methods. In U. Cress et al. (eds.), *International Handbook of Computer-Supported Collaborative Learning*, Computer-Supported Collaborative Learning Series 19, 63–83, Springer Nature Switzerland, https://doi.org/10.1007/978-3-030-65291-3_4

Järvelä, S., Kirschner, P. A., Hadwin, A., Järvenoja, H., Malmberg, J., Miller, M., & Laru, J. (2016). Socially shared regulation of learning in CSCL: Understanding and prompting individual- and group-level shared regulatory activities. *International Journal of Computer-Supported Collaborative Learning*, *11*(3), 263–280. https://doi.org/10.1007/s11412-016-9238-2

Järvelä, S., Nguyen, A., Vuorenmaa, E., Malmberg, J., & Järvenoja, H. (2023). Predicting regulatory activities for socially shared regulation to optimize collaborative learning. *Computers in Human Behavior, 144*, 107737. https://doi.org/10.1016/j.chb.2023.107737

*Kim, D., Majdara, A., & Olson, W. (2024). A Pilot Study Inquiring into the Impact of ChatGPT on Lab Report Writing in Introductory Engineering Labs. *International Journal of Technology in Education*, *7*(2), 259–289. https://doi.org/10.46328/ijte.691

Liu, L., Hao, J., von Davier, A. A., Kyllonen, P., & Zapata-Rivera, D. (2015). A tough nut to crack: Measuring collaborative problem solving. In Y. Rosen et al. (Eds.), *Handbook of Research on Technology Tools for Real-World Skill Development: IGI Global*. https://doi.org/10.4018/978-1-4666-9441-5

*Liu, Z., Zhu, Z., Zhu, L., Jiang, E., Hu, X., Peppler, K., & Ramani, K. (2024). ClassMeta: Designing Interactive Virtual Classmate to Promote VR Classroom Participation. *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3613904.3642947

*Nguyen, A., Hong, Y., Dang, B., & Huang, X. S. (2024). Human-AI collaboration patterns in AI-assisted academic writing. *Studies in Higher Education, 49*(5), 847–864. https://doi.org/10.1080/03075079.2024.2323593

Noroozi, O., Weinberger, A., Biemans, H. J. A., Mulder, M., & Chizari, M. (2013). Facilitating argumentative knowledge construction through a transactive discussion script in CSCL. *Computers & Education, 61*, 59–76, https://doi.org/10.1016/j.compedu.2012.08.013

Rosé, C., Wang, Y.-C., Cui, Y., Arguello, J., Stegmann, K., Weinberger, A., & Fischer, F. (2008). Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in computer-supported collaborative learning. *International Journal of Computer-Supported Collaborative Learning*, *3*(3), 237–271. https://doi.org/10.1007/s11412-007-9034-0

Schnaubert, L., & Bodemer, D. (2022). Group awareness and regulation in computer-supported collaborative learning. *International Journal of Computer-Supported Collaborative Learning, 17*(1), 11–38. https://doi.org/10.1007/s11412-022-09361-1

*Tai, T.-Y., & Chen, H. H.-J. (2024). Navigating elementary EFL speaking skills with generative AI chatbots: Exploring individual and paired interactions. *Computers and Education, 220*. https://doi.org/10.1016/j.compedu.2024.105112

*Tzirides, A. O., Zapata, G., Kastania, N. P., Saini, A. K., Castro, V., Ismael, S. A., You, Y. L., dos Santos, T. A., Searsmith, D., O'Brien, C., Cope, B., & Kalantzis, M. (2024). Combining human and artificial intelligence for enhanced AI literacy in higher education. *Computers and Education Open, 6*, 13. https://doi.org/10.1016/j.caeo.2024.100184

Vogel, F., Wecker, C., Kollar, I., & Fischer, F. (2017). Socio-Cognitive Scaffolding with Computer-Supported Collaboration Scripts: A Meta-Analysis. *Educational Psychology Review, 29*(3), 477–511. https://doi.org/10.1007/s10648-016-9361-7

Walker, E., Rummel, N., & Koedinger, K. R. (2014). Adaptive Intelligent Support to Improve Peer Tutoring in Algebra. *International Journal of Artificial Intelligence in Education, 24*(1), 33–61. https://doi.org/10.1007/s40593-013-0001-9

Weinberger, A., Ertl, B., Fischer, F. et al. Epistemic and social scripts in computer–supported collaborative learning. Instructional Science 33(1), 1–30 (2005). https://doi.org/10.1007/s11251-004-2322-4

Wise, A. F., & Schwarz, B. B. (2017). Visions of CSCL: Eight provocations for the future of the field. *International Journal of Computer-Supported Collaborative Learning, 12*(4), 423–467. https://doi.org/10.1007/s11412-017-9267-5

Yan, L., Greiff, S., Teuber, Z., & Gašević, D. (2024). Promises and challenges of generative artificial intelligence for human learning. *Nature Human Behaviour, 8*(10), 1839–1850. https://doi.org/10.1038/s41562-024-02004-5

References marked with an asterisk * indicate studies included in the review.

International Society of the Learning Sciences

Collaboration Quality in Computational Modeling: Implications for Middle-School STEM+C Learning

Shruti Jain, Namrata Srivastava, Srigowri Mayasandra Prasanna, Clayton Cohn, Umesh Timalsina, Naveeduddin Mohammed, Gautam Biswas
shruti.jain@vanderbilt.edu, Namrata.Srivastava@vanderbilt.edu, srigowri.mayasandra.prasanna@vanderbilt.edu, clayton.a.cohn@vanderbilt.edu, umesh.timalsina@vanderbilt.edu, naveeduddin.mohammed@vanderbilt.edu, gautam.biswas@vanderbilt.edu
Vanderbilt University, USA

Abstract: Collaboration is recognized by NGSS as an important 21st-century skill, yet its role in middle-school STEM+C remains underexplored. This study examines how students collaborate while building computational models (CM) of scientific processes and how it relates to learning outcomes. We coded 16 hours of activity from eight dyads across Meier et al.'s nine collaboration dimensions (e.g., mutual understanding, dialogue management, information pooling) and analyzed links between collaboration quality, model performance, and assessment scores. Findings highlight the dynamics of CM collaboration and the need for scaffolds that promote productive interaction and teach collaboration as a transferable skill.

Introduction and background

Computational modeling, where students create models of scientific processes by translating scientific knowledge into computational form, has been shown to foster STEM+C learning (Grover and Pea, 2013; Hutchins et al., 2020a). This requires students to identify variables and their relationships, which can be challenging for middle school students still building foundational knowledge (Basu et al., 2016). However, most studies focus on secondary and university students who already have STEM+C backgrounds (Ma et al., 2022).

Collaboration enables students to share information and develop a shared understanding (Snyder et al., 2025), aligning with the social-constructivist view of learning as interaction-driven rather than as individually acquired (Vygotsky, 1978). From a CSCL perspective, productive collaboration involves the co-construction of knowledge through discourse, coordination, and reciprocal engagement (Roschelle & Teasley, 1995). Thus, in computational modeling, learning depends not only on participation but also on the quality of interaction as the students negotiate ideas and build shared representations of scientific processes. The NGSS recognizes both collaboration and computational modeling as essential 21st-century skills (NGSS, 2013).

Several CSCL studies analyze collaborative discourse in middle-school STEM+C contexts. Hong et al. (2025) found that frequent idea sharing and negotiation were associated with greater gains. Studies focusing on computational modeling emphasize information pooling and shared understanding (Snyder et al., 2024; 2025), framing collaboration as a transferable skill (Cohn et al., 2025). Notably, these studies focus on older learners, and research on younger students remains underexplored (Ma et al., 2022). Primary-level studies using block-based coding (e.g., Ardito et al., 2020) focus on computing rather than modeling scientific processes, missing the affordances of collaboration in open-ended, interdisciplinary problem solving. Thus, a significant gap remains in understanding how middle-school students collaborate to co-construct knowledge as they jointly model computational representations of scientific processes.

In this paper, we analyze 16 hours of video and conversation from eight pairs of middle school students working together in a STEM+C learning environment. Our analysis is guided by a social-constructivist CSCL framing in which learning is mediated through discourse-based processes of shared meaning-making and co-construction (Vygotsky, 1978; Roschelle & Teasley, 1995). Consistent with this framing, we use Meier et al.'s (2007) Rating Scheme as an analytic framework to operationalize these interactional processes in students' collaborative work to address two exploratory research questions: (1) Which dimensions of collaboration characterize middle school students' work during science-based computational modeling tasks? (2) How does collaboration quality relate to learning outcomes, including modeling performance and assessment scores?

Methods

WRC curriculum unit

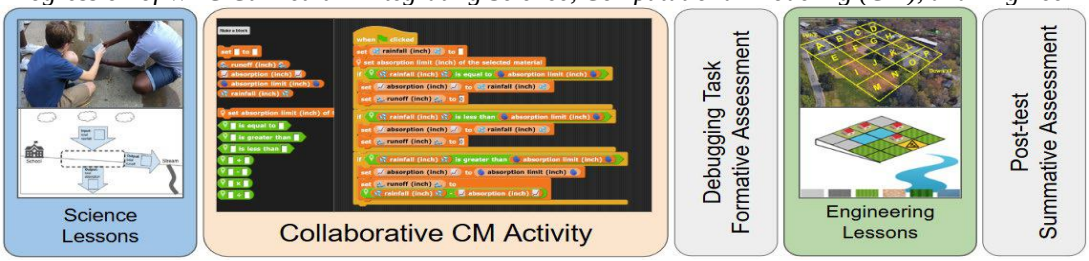
The WRC curriculum (Basu et al., 2022) comprises 15 lessons delivered over three weeks of one-hour in-person classes, integrating pen-and-paper activities with computer-based tasks. Focusing on Earth science concepts, such

 © ISLS

as conservation of mass, rainfall, absorption, and runoff, students examine the problems caused by urban water runoff. They perform hands-on investigations to explore how different surfaces absorb rainwater and produce runoff. Students develop conceptual models to calculate absorption and runoff, then implement them as computational models (CM). In the CM environment, teacher-assigned pairs build models using a block-based, domain-specific modeling language (DSML; Hutchins et al., 2020b). DSML combines domain components (rainfall, surface materials as inputs; absorption, runoff as outputs) with computational thinking practices (decomposition, algorithmic thinking) and computing concepts (variable initialization, conditional logic). After the CM activity, students undertake an engineering challenge to reduce schoolyard runoff within budget and accessibility constraints. Learning outcomes are evaluated through a post-test covering Science, Engineering, and Computational Thinking, complemented by formative assessments. Figure 1 provides an overview of the curriculum, highlighting key activities and the modeling environment.

Figure 1

Progression of WRC Curriculum Integrating Science, Computational Modeling (CM), and Engineering.



Participants and data

This study took place in a middle school (ages 10–12) during science lessons in the southeastern United States. A total of 104 students (51% male, 49% female; 67% White, 14% Black/African American, 11% Asian, 8% Hispanic/Latino) across four class periods participated, taught by two teachers. The study was approved by Vanderbilt University's IRB and the school district, with informed consent from parents and students. We collected students' assessment data, along with time-aligned audio, video, screen captures, and interaction logs from the CM environment. Instructional approaches differed: Teacher 1 (one class period) employed a scaffolded, step-by-step method; Teacher 2 (three class periods) used a flexible, student-driven approach. We focused on Teacher 2's classroom to explore authentic collaboration in open-ended problem-solving and its connection to learning outcomes. Absences during CM sessions or assessments led to incomplete datasets, resulting in a final sample of eight pairs, with two one-hour sessions, formative assessments, and post-tests.

Analysis methods

To analyze the video data, we segmented it using time-aligned interaction logs at each model test run. This yielded 80 segments (M=10, SD=5.2), lasting 15 seconds to 13.5 minutes (M=3:00, SD=2:56). Segments were deductively coded to map onto Meier et al.'s (2007) nine collaboration dimensions using the codebook described in Table 1. Meier et al.'s scheme was selected because it operationalizes collaboration as observable interactional processes, including shared understanding, reciprocity, coordination, and joint task engagement. These interaction processes can be linked to social constructivist theory constructs that emphasize the use of language (communication) as a mediating tool to develop shared understanding, co-construct knowledge, and coordinate shared tasks to facilitate collaborative learning. This makes the framework well-suited for examining how discourse quality mediates learning in collaborative computational modeling. The codebook was iteratively refined by two researchers, resolving discrepancies through discussion, establishing inter-rater reliability (Cohen's $\kappa = 0.95$). Each code was mapped to a discrete event, and a rationale was recorded. To answer RQ1, we conducted frequency analyses and computed descriptive statistics on the assigned codes to identify which aspects of collaboration were most prominent in students' conversations. For RQ2, we computed a *Collaboration Quality Score* by adding the percentage frequencies across collaboration dimensions, and analyzed its relation to students' CM, formative assessment (FA), and post-test CT scores:

1. *Computational Modeling (CM) Scores*: The CMs were evaluated for accuracy, with 1 point awarded for each correct line of code. Figure 1 (middle) shows the correct model, with a total score of 12 points.
2. *Debugging Task FA Scores*: Following the CM activity, students were assessed on their ability to interpret and predict model outputs by debugging a fictional student's erroneous model. A total of 9 points: two for identifying buggy outputs, two for corrected outputs, and one for each of the five errors.

3. *Post-Test CT scores:* At the end of the curriculum, the post-test assessed students' computational thinking skills (variables, conditional statements, control flow) framed in everyday contexts. Three post-test items, worth 13 points, were scored by two researchers (Cohen's $\kappa_w = 0.822$).

Given the small sample size and exploratory nature of the analysis, we used correlations and descriptive statistics to identify insights from student pair characteristics. The results are shown in the next section.

Table 1
Codebook for Evaluating Collaboration Quality Dimensions (Meier et al., 2007) in Students' Conversations.

Dimension	Theoretical Description	WRC Example
SMU	Moments when both partners actively ensure shared comprehension.	A partner checks other's understanding/explains their environment actions. E.g., "I think you are right."
DM	Clear turn-taking cues, mutual attention checks, avoiding confusion or interruptions.	Listen and wait for the other student to finish before speaking; Acknowledge. E.g., "What were you saying?"
IP	Participants introduce new relevant information.	Read from resources; seek help from a researcher or peers; recall previous activities. E.g., "Hey, rainfall will be less than the absorption limit"
RC	Critical discussions like evaluating alternatives, exchanging arguments, and presenting evidence.	Arguing about use of variables, conditional statements, or expressions. E.g., S1: "Set the rainfall value", S2: "No, we did that", S1: "okay, we go run the model"
TD	Planning/dividing tasks, drawing from expertise.	Assign tasks to each other or ask each other to run the test. E.g., "Click the green flag"
TM	Time management to ensure steady progress.	Check or verbalize the time remaining for the task. E.g., "24 minutes left…"
TC	Coordinate environment use via turn-taking.	Alternating computer usage during model construction. E.g., "Okay. Now you do it."
RI	Partners are respectful, encouraging the contribution of opinions.	Students consider each other's opinions, responding respectfully. E.g., "We did it!"
ITO	Captures motivation, including individual effort/engagement.	Working on task; minimal off-topic talk; regain focus after breaks. E.g., "Anyway, let's be serious…"

SMU: Sustaining Mutual Understanding, DM: Dialogue Management, IP: Information Pooling, RC: Reaching Consensus, TD: Task Division, TM: Time Management, TC: Technical Coordination, RI: Reciprocal Interaction, ITO: Individual Task Orientation (Meier, 2007)

Results

RQ1: Which dimensions of collaboration characterize middle school students' work during science-based computational modeling tasks?

Frequency analysis of the 80 coded segments revealed distinct patterns in students' collaborative behaviors during the computational modeling activity (RQ1). Students exhibited strong task-oriented behavior (see full table in Supplementary Materials (1)), with *Individual Task Orientation* observed in 81.3% and 92.5% for Student 1 and Student 2, respectively. A substantial proportion of student dialogue involved *Information Pooling* (30%), particularly during the initial stages of model construction, as students frequently consulted their notebooks for instructions to guide their work. Evidence of *Sustaining Mutual Understanding appeared in approximately 25%* of our dataset, reflecting efforts to maintain shared meaning. In contrast, *Dialogue Management* (10%) and *Reaching Consensus* (8.75%), which are central to productive collaboration, were relatively infrequent. *Task Division* was also rare (11.25%), as students were encouraged to build the computational model together. However, students frequently alternated control of the laptop, reflecting *Technical Coordination* (20%). *Time Management* was also seldom observed (6.25%), perhaps because the teacher and researchers externally regulated pacing. The codes *Dialogue Management, Reaching Consensus, Task Division,* and *Time Management* each appeared in fewer than 10 segments; hence, they were excluded from subsequent analysis.

RQ2: How does collaboration quality relate to individual learning outcomes?

To examine how collaboration quality relates to learning outcomes, we calculated Pearson's correlations between the collaboration dimensions (*Sustaining Mutual Understanding*, *Information Pooling*, *Technical Coordination*, *Reciprocal Interaction*, and *Individual Task Orientation*) and the three performance measures: CM score, debugging task score, and post-test CT score. See details in Supplementary Materials (1).

 © ISLS

Among pairs that completed the CM, at least one partner was always engaged (*Individual Task Orientation* = 100%). The average *Individual Task Orientation* was strongly correlated with CM score: *r=0.69*. *Sustaining Mutual Understanding* and *Reciprocal Interaction* were also positively associated with CM score (*r=0.43* and *r=0.34*, respectively), whereas *Information Pooling (r=0.13)* and *Technical Coordination* (*r=-0.06)* had minimal influence. The aggregated Collaboration Quality Score (see Methods) was moderately correlated with CM score (*r=0.42)*, highlighting the role of collaborative processes in science-based modeling tasks. Good collaboration was also conducive to promoting students' individual learning, both in the debugging task (*r=0.58*), reflecting a more immediate, domain-proximal association, and in the post-test (*r=0.54*), reflecting a longer-term association. *Sustaining Mutual Understanding* was strongly correlated with average debugging score (*r=0.73*) and post-test (*r=0.59*). Similarly, *Information Pooling* (*r=0.49* with debugging*; r=0.51* with post-test*) and Reciprocal Interaction* (*r=0.58* with debugging*; r=0.50* with post-test) were also positively associated. In contrast, *Task Orientation* showed only a moderate association with assessments: *r=0.34* for debugging and *r=0.24* for the post-test. These findings reinforce that shared sense-making processes: *Sustaining Mutual Understanding*, *Information Pooling*, and *Reciprocal Interaction*, are tightly coupled with conceptual learning in collaborative computational modeling activity.

Case study: The Podcast Boys

This case study examines a pair of students, "Podcast Boys", an intriguing instance where low collaboration quality coincided with notable immediate, domain-proximal learning, but limited transfer to post-test questions. During the CM activity, the students humorously pretended to record a podcast of their work, speaking aloud to an "audience" rather than engaging in *Reciprocal Interaction*. Only S2 was engaged throughout (*Individual Task Orientation = 100%*) and completed the model alone (*Technical Coordination = 0%*). The brief instances of *Sustaining Mutual Understanding* occurred only at the beginning and end of the activity, both times when the students explained their progress to the imagined audience. As such, both students performed equally on the Debugging Task (6 points each), suggesting a shared understanding of the final model; however, this did not transfer, as S2 scored 9 points on the post-test, while S1 scored only 3 (see our Supplementary Materials (1)).

While the students' "podcast" approach fostered continuous verbalization and externalization—a practice that can support metacognitive awareness (Schraw & Dennison, 1994)—it fell short of genuine collaboration, which requires reciprocal contributions, negotiation of meaning, and shared problem-solving (Roschelle & Teasley, 1995). This case highlights that externalization alone is insufficient and that effective collaboration requires mutual engagement and the co-construction of understanding.

Discussion and conclusion

Our findings extend the growing body of research on collaboration by illuminating how middle school students engage in and benefit from collaborative practices during science-based modeling tasks. Theoretically, this study deepens social constructivist accounts of CSCL by illustrating how collaboration quality is manifested through observable discourse processes rather than participation frequency. The prominence of *Sustaining Mutual Understanding*, *Information Pooling*, and *Reciprocal Interaction*, and their strong associations with learning outcomes, reinforces the idea that conceptual learning in computational modeling emerges through socially mediated sense-making. Our results emphasize that effective collaboration extends beyond technical coordination and task division; it involves reciprocity and the co-construction of ideas (Roschelle & Teasley, 1995; Snyder et al., 2025). The relative absence of higher-order collaborative behaviors—such as *Reaching Consensus* and *Dialogue Management*—highlights that at the middle school level, collaboration skills require explicit cultivation (Barron, 2003). This calls for designing explicit scaffolds for promoting collaboration in the learning environment through teacher prompts, interface features, or task structures.

This study is not without limitations. The multimodal data collection enabled fine-grained analysis, but introduced data loss: missing streams or student absences during the longitudinal sequence of lessons or assessments, rendering the dyad's data unusable. Consequently, only eight dyads were retained. Our findings are also correlational and exploratory, aimed at characterizing collaborative processes and their relationship with learning outcomes. However, our preliminary results point to embedding structured opportunities for students to practice and refine their collaborative strategies within computational modeling tasks to support STEM+C learning, with collaboration serving as both a process and an outcome of learning in technology-enhanced learning environments. Furthermore, with recent advances in LLMs and generative AI, we can develop agents to support the development of effective collaborative processes that, in turn, scaffold learning.

Endnotes

(1) https://docs.google.com/document/d/1IAmiIbY-L1diM6vsVDBzAFE3u-nA0ONE/edit

References

Ardito, G., Czerkawski, B., & Scollins, L. (2020). Learning computational thinking together: Effects of gender differences in a collaborative middle school robotics program. TechTrends, 64(3), 373-387.

Barron, B. (2003). When smart groups fail. The Journal of the Learning Sciences, 12(3), 307–359.

Basu, S., Biswas, G., Sengupta, P., Dickes, A., Kinnebrew, J. S., & Clark, D. (2016). Identifying middle school students' challenges in computational thinking-based science learning. Research and practice in technology-enhanced learning, 11(1), 13.

Basu, S., McElhaney, K. W., Rachmatullah, A., Hutchins, N. M., Biswas, G., & Chiu, J. (2022). Promoting computational thinking through science-engineering integration using computational modeling. In Proceedings of the 16th International Conference of the Learning Sciences-ICLS 2022, pp. 743-750. International Society of the Learning Sciences.

Cohn, C., Fonteles, J. H., Snyder, C., Srivastava, N., Campbell, D., Montenegro, J., & Biswas, G. (2025). Exploring the design of pedagogical agent roles in collaborative stem+ c learning. In Proceedings of the 18th International Conference on Computer-Supported Collaborative Learning-CSCL 2025, pp. 330-334. International Society of the Learning Sciences.

Grover, S., & Pea, R. (2013). Computational thinking in K–12: A review of the state of the field. Educational researcher, 42(1), 38-43.

Hutchins, N. M., Biswas, G., Maróti, M., Lédeczi, Á., Grover, S., Wolf, R., ... & McElhaney, K. (2020a). C2STEM: A system for synergistic learning of physics and computational thinking. Journal of Science Education and Technology, 29(1), 83-100.

Hutchins, N. M., Biswas, G., Zhang, N., Snyder, C., Lédeczi, Á., & Maróti, M. (2020b). Domain-specific modeling languages in computer-based learning environments: A systematic approach to support science learning through computational modeling. International Journal of Artificial Intelligence in Education, 30, 537-580.

Hong, D., Feng, C., Zou, X., Huang, S., Hmelo-Silver, C. E., Wang, T., ... & Lester, J. C. (2025). A Clustering Approach to Collaborative Problem Solving: Utilizing Discourse and Trace Log Data. In Proceedings of the 18th International Conference on Computer-Supported Collaborative Learning-CSCL 2025, pp. 81-89. International Society of the Learning Sciences.

Ma, Y., Martinez Ruiz, J., Brown, T. D., Diaz, K. A., et al. (2022, February). It's Challenging but Doable: Lessons Learned from a Remote Collaborative Coding Camp for Elementary Students. In Proceedings of the 53rd ACM Technical Symposium on Computer Science Education-Volume 1 (pp. 342-348).

Meier, A., Spada, H., & Rummel, N. (2007). A rating scheme for assessing the quality of computer-supported collaboration processes. International Journal of Computer-Supported Collaborative Learning.

NGSS Lead States. (2013). Next generation science standards: For states, by states. National Academies Press.

Roschelle, J., & Teasley, S. D. (1995, August). The construction of shared knowledge in collaborative problem solving. In Computer-supported collaborative learning (pp. 69-97). Springer Berlin Heidelberg.

Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. Contemporary educational psychology, 19(4), 460-475.

Snyder, C., Wen, C. T., Hutchins, N. M., Vatral, C., Liu, C. C., & Biswas, G. (2024). Investigating collaborative problem-solving behaviors during STEM+C learning in groups with different prior knowledge distributions. In Proceedings of the 17th International Conference on Computer-Supported Collaborative Learning-CSCL 2024, pp. 107-114. International Society of the Learning Sciences.

Snyder, C., Cohn, C., Fonteles, J. H., & Biswas, G. (2025). Using collaborative interactivity metrics to analyze students' problem-solving behaviors during STEM+C computational modeling tasks. Learning and Individual Differences, 121, 102724.

Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes (Vol. 86). Harvard University Press.

Acknowledgment

This work is supported by the National Science Foundation under award DRL-1742195. The opinions expressed are those of the authors and do not represent the views of the National Science Foundation.

International Society of the Learning Sciences

Scaffolding Knowledge Building through Ordered-Epistemic Network Analysis-Supported Reflective Assessment and Meta-Discourse

Xiaojie Zhang, The University of Hong Kong, zxjie@connect.hku.hk,
Carol Chan, The University of Hong Kong, ckkchan@hku.hk
Shihui Feng, The University of Hong Kong, shihuife@hku.hk
Jianhua Zhao, Southern University of Science and Technology, zhaojh@sustech.edu.cn

Abstract: This study presents a reflective assessment approach supported by ordered-epistemic network analysis (ONA) tools, designed to enhance students' knowledge building (KB) and discourse within the Knowledge Forum. While prior research has demonstrated that reflective assessments integrated with learning analytics can improve students' metacognitive skills and self-regulation in knowledge-building contexts, this study uniquely incorporates ONA-based reflection and meta-discourse to elucidate and promote the dynamics and trajectories of epistemic moves during the knowledge-building process. The participants comprised 40 Year 5 students engaged in a collective inquiry into metabolism using the Knowledge Forum. Quantitative results indicated that, following the intervention, students exhibited increased levels of social and community interaction, alongside more sophisticated epistemological beliefs. Qualitative analysis further illustrated how ONA-facilitated reflective assessment enriched students' collaborative inquiry process. This work advances pedagogy in knowledge building by demonstrating that ONA-enhanced reflective assessment and meta-discourse can effectively support and deepen collaborative knowledge construction.

Introduction

Knowledge building (KB) is a collaborative learning process within a community, emphasising the continuous development and refinement of ideas and theories (Scardamalia & Bereiter, 2010). It centres on collective knowledge creation to enhance shared understanding. Knowledge Forum® (KF) is a computer-mediated platform designed to facilitate communal knowledge advancement through epistemic activities such as problem formulation, idea elaboration, explanation co-construction, referencing authoritative sources, theory revision, and reflective synthesis. Research shows that explanation-driven inquiry underpins scientific communities, with questions and explanations progressively advancing knowledge (Zhang et al., 2007). Sustained questioning, idea improvement, and synthesis of diverse ideas are key for fostering advanced knowledge (Chan et al., 2018). While KF supports ongoing discourse and explanatory inquiry, a challenge lies in students' difficulty tracking different discussion threads. Scaffolding discourse remains vital, with approaches including teacher-led knowledge-building talks, reflective portfolios (Van Aalst & Chan, 2007), and metacognitive conversations (Zhang et al., 2015). Recent methods utilise learning analytics such as word clouds (Resendes et al., 2015), knowledge connection analysers (Yang et al., 2021), and KBDeX (Oshima et al., 2012) to visualise idea progression and support reflection (Yang et al., 2022). However, research on students' reflections on their epistemic contributions, especially their dynamics and trajectories, is limited.

Ordered-epistemic network analysis (ONA) offers a promising approach by revealing the temporal sequences of epistemic activities, enabling both quantitative and qualitative analysis of collaborative interactions (Tan et al., 2020). While ONA has been utilised to analyse epistemic moves within KB (Yang et al., 2022), few studies have explored its pedagogical applications. Our prior work demonstrated ONA's potential in scaffolding reflective assessment and improving domain understanding (Zhang et al., 2025). This study further investigates classroom meta-discourse, ONA-reflective assessment, and their effects on social interaction and epistemic beliefs (EBs). The research questions are: RQ1: Does ONA-enhanced reflective assessment support improvements in social interactions and epistemological beliefs? RQ2: How do students engage with ONA in reflective assessment to advance knowledge?

Method

Research participants and context

 © ISLS

This study was conducted at a primary school in Shenzhen, China, involving 40 fifth-grade students engaged in a science inquiry on human metabolism. The inquiry activities took place on Knowledge Forum® (KF; https://hku.knowledgeforum.org), a platform commonly used in knowledge-building research (Chen & Hong, 2016) to facilitate the collective construction of knowledge (Scardamalia & Bereiter, 2014).

ONA visualisations for reflective assessment

To facilitate students' reflection, the ONA visualisation for students was constructed. First, we coded students' KF discourse (see Table 1). Twenty per cent of notes were independently coded by two researchers in CSCL, with Cohen's kappa = 0.830. Second, we conducted ONA analyses to depict the temporal transitions among the codes for different selected sub-topics/discourse threads (see Figure 1).

During the reflective assessment, each student was provided with an ONA diagram for a specific sub-topic and asked to respond to two prompts: "*What did I notice from the ONA diagram?" and "How can we further develop our knowledge on KF?"*

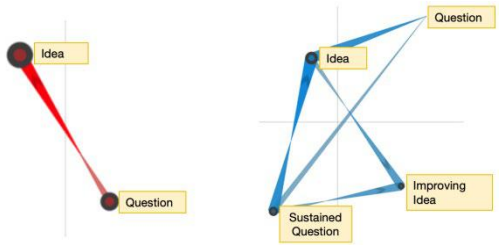
Table 1

The Coding Scheme of KF Discourse is Based on Discourse Moves

Epistemic Move	Definition
Initial Question	A student asks questions seeking elaborative explanations.
Sustained Question	A student asks further questions about an existing idea.
Idea	A student proposes an idea.
Improving Idea	A student improves an already existing idea by providing deeper or broader perspectives.
Explanation	Make one's viewpoint easier to understand.
Extraneous Information	A student posts some extraneous information from Baidu or LLM.
Synthesizing	Extract or summarize the previous diverse discussion content.

Figure 1

Two Examples of the ONA Diagrams



Pedagogical design

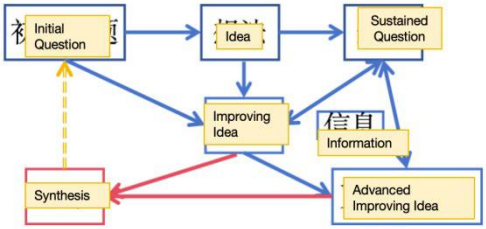
The course lasted for 3 weeks, including 8 lessons (each lasting 45 minutes). The first 5 lessons focused on introducing and engaging students in KB, and the last 3 emphasised ONA-supported reflective assessment designs. These three lessons aim to enable students to recognise and develop a deeper understanding of the collective knowledge development trajectory supported by classroom meta-discourse and ONA visualisation for more active KB work. We developed these lessons in progressive ways for illuminating the KB trajectory, which includes sustained questioning of existing ideas, improving ideas with information and explanations, and synthesising diverse perspectives using meta-discourse and ONA reflection.

Each ONA-supported lesson comprised two main components: (1) classroom meta-discourse on KF discourse to help students understand the KB trajectories (Figure 2), and (2) a reflective assessment activity guided by ONA visualisation (Figure 1) to help students reflect and modify the trajectory in KF discourse. Students continued to contribute to the KF after each lesson. The first lesson focused on changing the basic trajectory from "Initial Question → Idea" into "Initial Question → Idea → Sustained Question". The second lesson extended this to "Initial Question → Idea → Sustained Question ↔ Improving Idea". The third lesson further developed the trajectory to "Initial Question → Idea → Sustained Question ↔ Improving Idea (with Information) → Synthesis".

Figure 2

The Progressively Sophisticated KB Trajectory

 © ISLS



Data collection and analysis

Social Interaction: Social Network Analysis. The "who built on whose notes" matrices derived from the KF log data were transformed into social networks to characterise students' social interactions. In this study, weighted out-degree means how many times one student built on others' notes. We used the percentage of weighted out-degrees to characterise the extent to which the students engage in collective knowledge creation. Paired sample t-tests were employed to compare those SNA indices between the experimental and comparison classes.

Epistemological Beliefs: pre/post-test. Each student was asked to respond to the following questions: "Where do I think ideas come from?" and "Why do we need ideas? What can ideas do?" Their responses were coded according to Popper's framework (Hong & Chiu, 2016; Popper, 1979). Responses that described beliefs as mental representations within individual minds were classified as World 2, whereas those viewing ideas as collective creations emerging from group discussions were categorised as World 3 (Zhang et al., 2026). All students' writings on EBs were independently coded by one researcher and DeepSeek. The Cohen's kappa coefficients for the two questions were 0.825 and 0.771, respectively. The teacher was blind to the students' EB scores, which helped minimise potential bias or influence related to these measures.

Student engagement in analytics-supported knowledge building. Students' responses to the worksheets are analysed to examine their engagement with the ONA-supported reflective assessment. As well, an interview was conducted with several students. The interview transcript and reflection worksheets were collected to analyse their engagement in ONA-supported reflection activities. A qualitative analysis was conducted to examine how students engaged in the ONA-supported reflective assessment and KB inquiry.

Results

RQ1: Did the ONA-enhanced reflective assessment support students' improvement of epistemological beliefs?

Social interactions

The results of the percentage of weighted out-degree and the paired sample t-test are presented in Table 4. The paired sample t-test revealed that, after lesson 1 and lesson 2, the percentage of weighted out-degree increased significantly (t = -4.045, $p < 0.001$; t = -3.784, p = 0.001). These findings indicate that after engaging in meta-discourse and reflective assessment focused on "sustained questions" and "improving ideas," students were more likely to build on existing ideas by proposing sustained questions and improving ideas.

Epistemological beliefs

For the question "Where do I think 'ideas' come from?", more students adopted a World 3 perspective after the intervention (24/40, 60%) compared to before the intervention (16/40, 40%). However, the chi-square test did not indicate statistical significance (χ^2 (1, n = 80) = 3.20, p = 0.074). For the question "Why do we need 'ideas? What can 'ideas' do?", the number of students holding a World 3 perspective increased from 17/40 (42.5%) before the intervention to 30/40 (75%) afterwards, and this difference was statistically significant (χ^2 (1, n = 80) =8.76, p = 0.003). Overall, the intervention enhanced students' EBs, shifting from a World 2 to a World 3 perspective. For example, one student wrote before the intervention, "Ideas come from our mind; ideas are used to think," but by the end of the course, stated, "Ideas come from questions; we need ideas because they can generate new questions; ideas enable us to solve problems."

RQ2: How did the students engage in knowledge advances and ONA-enhanced reflective assessment?

We presented the key events from students' KB talk in the classroom, students' responses to ONA worksheets,

 © ISLS

and interviews to analyse student engagement in community knowledge advancement and in ONA-supported reflective activities are also examined.

Theme 1: Posting questions and ideas. Before the intervention, students posted their initial questions and ideas about metabolism on the KF view. Their questions included: "What does metabolism include?" "What is the relationship between metabolism and exercise?" "What is the link between metabolism and obesity?" "Why does metabolism weaken with age?" "How is metabolism related to height?", and "What are the advantages and disadvantages of slow metabolism?"

Theme 2: Asking sustained questions. In the reflective ONA worksheets from lesson 1, most students recognised the importance of sustained questioning. They noted, "After the classroom discussion, we found that the KB trajectory 'Question → Idea → Sustained Question' can promote knowledge creation. In our previous KF discourse, we lacked sustained questions." Subsequently, students began to ask sustained questions to deepen existing ideas in KF. For example, in response to "What is the relationship between metabolism and exercise?", one student answered, "High-intensity exercise accelerates metabolism," while another sustained with, "Why does exercise accelerate metabolism?"

Theme 3: Improving existing ideas. In the reflective ONA worksheets from lesson 2, students observed, "Sustained questions arise from ideas, and ideas come from initial questions." They also commented, "The line between sustained questions and improving ideas is subtle," and "Some ideas are too simple; we need more improved ideas." As a result, students sought additional information from the internet or books and proposed improving ideas in response to sustained questions on KF. For instance, one student wrote: "During exercise, our bodies require a large amount of oxygen, which enters the bloodstream and circulates through the heart. This process involves metabolism, as the heart helps remove waste products like carbon dioxide, thereby accelerating blood circulation and expelling these wastes from the body."

Theme 4: Synthesising diverse ideas to answer initial questions. In the reflective ONA worksheets from lesson 3, students wrote: "From the ONA diagram, I noticed there are many connections between sustained questions and ideas, and more between sustained questions and improving ideas. We need to do more summarising of ideas and explanations, as this helps make complex thoughts more concise and understandable." Each student selected an initial question and wrote a synthesis of improving ideas related to that question.

Discussion and conclusion

This study investigated how ONA-enhanced reflective assessment can support students in reflecting on their KB trajectories on KF and in adjusting their actions during subsequent phases of inquiry. The quantitative findings indicate that students shifted from a World 2 perspective (focused on personal and subjective understanding) to a World 3 perspective (centred on collective, objective knowledge), alongside demonstrating increased levels of community-collaborative interaction. These results align with our initial conjecture: through developing a deeper understanding of KB trajectories during the ONA-enhanced reflective assessment process, students gained insights into the nature of collective knowledge with richer social interactions in KF discourse. The ONA analysis also suggests that students with higher ONA-reflection scores, demonstrating better understanding of ONA visualisation and reflection of the KB learning trajectory, demonstrate better epistemic action patterns.

The qualitative analysis provides further nuance, showing that students actively examined their ONA diagrams to identify areas for improvement and to strategically refine their KF discourse. Students developed more understanding, progressing from a relatively basic trajectory ("Initial Question → Idea") to a more sophisticated pattern ("Initial Question → Idea → Sustained Question ↔ Improving Idea (with Information) → Synthesis"). This shift reflects not only an increased awareness of KB trajectories but also a growing recognition of the temporal and iterative nature of the KB process. By making discourse patterns visible, ONA visualisations and reflection prompts empower students to engage in the KB trajectory.

These findings align with and extend prior research on using reflective assessment (Resendes et al., 2015; Yang et al., 2021) for advancing knowledge building. We pioneered the use of ONA visualisation-reflection for school-aged children, and our initial work has demonstrated that ONA interventions can enhance students' domain understanding (Zhang et al., 2025). The present study goes further by examining the effects of ONA-based reflective assessment on students' epistemological beliefs and social interactions, and connecting the more closely related relations of ONA reflection with epistemic action on the Knowledge Forum. This research contributes to the advancement of KB pedagogy by showing that reflective assessment of epistemic participation, when supported by ONA visualisations, fosters more effective collaborative knowledge construction.

References

Bruning, R. H., Schraw, G. J., & Ronning, R. R. (1999). Cognitive psychology and instruction, perception of classroom environment, achievement goals, and achievement outcomes. *Journal of Educational Psychology, 93*(1), 43–54.

Chan, C. K., & Van Aalst, J. (2018). Knowledge building: Theory, design, and analysis. In *International handbook of the learning sciences* (pp. 295-307). Routledge.

Chen, B., & Hong, H. Y. (2016). Schools as knowledge-building organisations: Thirty years of design research. *Educational psychologist, 51*(2), 266-288.

Knoke, D., & Yang, S. (2020). *Social Network Analysis*. SAGE Publications, Inc.

Kuhn, D., Cheney, R., & Weinstock, M. (2000). The development of epistemological understanding. *Cognitive Development, 15*(3), 309–328.

Hong, H.-Y., & Chiu, C.-H. (2016). Understanding how students perceive the role of ideas for their knowledge work in a knowledge-building environment. *Australasian Journal of Educational Technology, 32*(1), Article 1.

Oshima, J., Oshima, R., & Matsuzawa, Y. (2012). Knowledge building discourse explorer: A social network analysis application for knowledge building discourse. *Educational technology research and development, 60*(5), 903-921.

Ouyang, F. (2021). Using Three Social Network Analysis Approaches to Understand Computer-Supported Collaborative Learning. *Journal of Educational Computing Research*, *59*(7), 1401–1424.

Popper, K. R. (1979). *Objective knowledge: An evolutionary approach*. Oxford, England: Clarendon Press.

Resendes, M., Scardamalia, M., Bereiter, C., Chen, B., & Halewood, C. (2015). Group-level formative feedback and metadiscourse. *International Journal of Computer-Supported Collaborative Learning, 10*(3), 309-336.

Scardamalia, M., & Bereiter, C. (2010). A brief history of knowledge building. *Canadian Journal of Learning and Technology/La revue canadienne de l'apprentissage et de la technologie, 36*(1).

Scardamalia, M., & Bereiter, C. (2014). Knowledge Building and Knowledge Creation: Theory, Pedagogy, and Technology. In R. K. Sawyer, *Cambridge handbook of the learning sciences* (2nd ed., pp. 397–417). Cambridge University Press.

Tan, Y., Ruis, A. R., Marquart, C., Cai, Z., Knowles, M. A., & Shaffer, D. W. (2022, October). Ordered network analysis. In *International Conference on Quantitative Ethnography* (pp. 101-116). Cham: Springer Nature Switzerland.

Van Aalst, J., & Chan, C. K. (2007). Student-directed assessment of knowledge building using electronic portfolios. *The Journal of the Learning Science*s, 16(2), 175-220.

Yang, Y., van Aalst, J., & Chan, C. (2021). Examining online discourse using the knowledge connection analyser framework and collaborative tools in knowledge building. *Sustainability*, 13(14), 8045.

Yang, Y., Yuan, K., Feng, X., Li, X., & van Aalst, J. (2022). Fostering low-achieving students' productive disciplinary engagement through knowledge-building inquiry and reflective assessment. *British Journal of Educational Technology, 53*(6), 1511-1529.

Zhang, J., Scardamalia, M., Lamon, M., Messina, R., & Reeve, R. (2007). Socio-cognitive dynamics of knowledge building in the work of 9-and 10-year-olds. *Educational Technology Research and Development, 55*(2), 117-145.

Zhang, J., Scardamalia, M., Reeve, R., & Messina, R. (2009). Designs for collective cognitive responsibility in knowledge-building communities. *The Journal of the Learning Sciences, 18*(1), 7-44.

Zhang, J., Tao, D., Sun, Y., Chen, M. H., Peebles, B., & Naqvi, S. (2015). Metadiscourse on collective knowledge progresses to inform sustained knowledge-building discourse. In the *Annual Meeting of American Educational Research Association, Chicago, IL*.

Zhang, X., Feng, S., Zhao, J., Zhang, Y., Yu, Y., & Chan, C. (2026). Relationships of students' epistemological beliefs with social interactions and epistemic actions in collaborative inquiry learning. *Interactive Learning Environments*, 1-18.

Zhang, X., Chan, C., Feng, S., Ouyang, W., Xie, Y., Zhao, J., & Feng, X. (2025). Scaffolding Knowledge Building Processes through ONA-Supported Collective Reflective Assessment. In *Proceedings of the 18th International Conference on Computer-Supported Collaborative Learning-CSCL 2025*, pp. 214-222. International Society of the Learning Sciences.

Zhu, G., Scardamalia, M., Moreno, M., Martins, M., Nazeem, R., & Lai, Z. (2022). Discourse moves and emotion in knowledge-building discourse and metadiscourse. *In Frontiers in Education* (Vol. 7, p. 900440). Frontiers Media SA.

International Society of the Learning Sciences

Facilitating High School Students' Socially Shared Regulation of Learning Through STEM Inquiry into Local Air Quality and Asthma

Ho Sue Kwon, Catherine Roeroe, Susan A. Yoon
suekwon@upenn.edu, croeroe@upenn.edu, yoonsa@upenn.edu
University of Pennsylvania

Abstract: This exploratory study examines how collaborative learning environments mediated by mobile technology facilitate socially shared regulation of learning (SSRL) in high school STEM education. Ten high school students in three small groups investigated local air quality and asthma using handheld sensors to collect indoor and outdoor data. We analyzed 3.5 hours of video data using interaction and thematic analysis guided by the cognitive dimensions of SSRL. Findings show that SSRL emerged most prominently in moments of uncertainty which catalyzed collective monitoring, hypothesis generation, and goal-directed planning. Three design features supported these processes: (1) structured autonomy that shifted troubleshooting from external support to student-led, (2) authentic inquiry that turned technological breakdowns into regulatory opportunities, and (3) real-world constraints that extended collaboration. The study contributes to understanding SSRL as a naturally occurring phenomenon in authentic high school STEM contexts and offers design principles for leveraging uncertainty for facilitating SSRL.

Introduction

K-12 STEM education increasingly calls for collaborative, interdisciplinary inquiry that connects school science to real community problems. Bioinformatics, an emerging STEM field, offers such a context by linking genetics, environment, and health in ways that position biology as a living, socially relevant discipline (Yoon et al., 2025). In many urban districts, asthma and poor air quality disproportionately affect students and families and making local air quality investigation a compelling and personally meaningful topic for STEM learning. To capitalize on this kind of socio-scientific issue, students must not only engage with complex real-world content but also to regulate their own learning individually and collectively (Michalsky & Cohen, 2021).

Research on socially shared regulation of learning (SSRL) demonstrates that groups can regulate behavior, cognition, and affect at the collective level (Hadwin et al., 2018). However, much of this work has examined SSRL under conditions of external support or interventions such as structured classroom/online learning environments, digital tools including AI, or direct teacher facilitation (Järvenoja et al., 2020; Kim et al., 2025; Michalsky & Cohen, 2021; Panadero & Järvelä, 2015; Sharma et al., 2024; Sulla et al., 2023). Less is known about how SSRL emerges organically in student-led, real-world inquiry, where newness, uncertainty, and collaboration are inherent.

This study addresses that gap by examining how SSRL emerged naturally during place-based fieldwork and identifying the design features that elicited it without over-structuring collaboration. Through thematic analysis of small-group activity, the study examines how SSRL emerged and what design features facilitated it in an interdisciplinary high school STEM project on local air quality and asthma. In this project, students in small groups used mobile sensors to engage in first-hand empirical data collection and scientific reasoning about socio-scientific issues directly relevant to their lives. Our contribution is twofold: theoretically, we elaborate SSRL as a naturally occurring phenomenon most visible in moments of uncertainty; practically, we offer design principles for creating collaborative STEM environments that use uncertainty as a pedagogical resource. To explore these dynamics, we ask, How did the STEM learning environment support students' SSRL?

Background

SSRL and uncertainty

Meaningful collaborative learning requires students to strategically regulate their collective behavior, cognition, emotion, and motivation to meet shared goals. SSRL is a qualitatively different, group-level phenomenon that extends regulation from the "I" to the "we" level emerging through transactive exchanges where members negotiate shared goals, monitor collective progress, and co-construct understanding (Hadwin et al., 2018; Isohätälä et al., 2017; Järvenoja et al., 2020). Increased participation, well-rounded contribution patterns, and transactive talk are key empirical markers of its emergence (Isohätälä et al., 2017; Törmänen et al., 2022).

 © ISLS

A growing body of literature positions uncertainty as central to this collaborative regulatory process. Chen et al. (2024) distinguish between two forms of scientific uncertainty: conceptual uncertainty, related to gaps in content knowledge, and epistemic uncertainty, defined as the subjective experience of being unsure about how knowledge is generated. Epistemic uncertainty, in particular, drives students to generate tentative hypotheses, claims, and evidence as they work to deepen understanding. In collaborative group settings, learners' success at social regulation is shaped by how well they collectively manage such uncertainty (Dragnić-Cindrić et al., 2020). Schnaubert (2023) extends this by theorizing uncertainty as a metacognitive experience that links cognitive and emotional regulation: when learners make their uncertainty legible to peers, it prompts others to offer support, realign strategies, or deepen engagement. Manz et al. (2025) further argue that when students engage with and seek to resolve uncertainty, they develop deeper conceptual understanding and take greater ownership of their scientific thinking and decision-making. While these studies establish the relationship between uncertainty and regulation in online and classroom-based settings, less is known about how these dynamics unfold in field-based inquiry where uncertainty arises from real-world constraints such as sensor malfunctions, missing data, and unpredictable environments.

High school STEM inquiry and CSCL as a design context

This study situates SSRL within an interdisciplinary, socio-scientific curriculum focused on asthma and air quality. The curriculum integrated genetics, environmental science, and biology across a 20-lesson unit in which students used mobile sensors to collect, analyze, and compare, air quality data across indoor, outdoor, and national contexts (Yoon et al., 2022). This problem-based design engaged students in authentic scientific practices while connecting inquiry to real civic and health concerns in their communities.

The study draws on Computer-Supported Collaborative Learning (CSCL) to examine how technology-mediated environments foster social regulation and meaning-making. Mobile air-quality sensors and data visualization tools functioned not simply as instruments for measurement but as mediating artifacts that encouraged collaboration and supported students in externalizing reasoning, coordinating actions, and constructing shared understandings of real-world phenomena (Jeong et al., 2019). Episodes of technological breakdown and unexpected data created moments of uncertainty that invited groups to negotiate strategies, share responsibility, and sustain engagement. This study contributes to CSCL by demonstrating how tool-mediated, place-based inquiry can be designed to transform uncertainty into a catalyst for SSRL, and how technology and task structure can be orchestrated to cultivate collaboration in high school STEM contexts.

Methodology

Context, participants, and data sources

The study involved ten high school students, organized into three small groups, participating in a bioinformatics project designed to investigate the high rates of asthma in urban areas caused by poor air quality. Groups were formed by gender and grade level: one group of four 9th grade females, one group of three 11th grade females, and one group of three 12th grade males. The students were enrolled in three different public high schools in the large metropolitan area in the Northeast, and each group was taught by a different teacher. The teachers of the participants were recruited as a small, highly motivated sample with strong interest in the project which aligns with recommendations for early-stage exploratory design work (IES & NSF, 2013).

The analysis drew from *Lesson 11: Data Collection Around the School* which introduced students to the practice of conducting field-based investigations with mobile air quality sensors. The lesson was designed to help students recognize that scientific investigation requires not only disciplinary knowledge but also the development of practical skills and collaboration. The intended learning outcome was to learn how to plan and carry out investigations as a group while using technology to collect physical data outside of the classroom. These activities positioned students to participate in authentic scientific practices including generating, analyzing, and interpreting data while working collaboratively to address a real-world problem.

The data consisted of 3.5 hours of video recordings capturing students' use of mobile air quality sensors to measure carbon monoxide and fine particulate matter at both indoor and outdoor sites around their school buildings. Student groups began by planning and coordinating routes for data collection, then proceeded to gather and archive sensor readings in real time. Along the way, they troubleshot device malfunctions and negotiated procedural decisions, such as whether to substitute or revisit locations when data were missing. The classroom phase captured students as they organized, compared, and interpreted their findings by situating their local data within broader environmental concerns. Throughout these activities, students expressed and regulated a range of emotions, from excitement to curiosity to frustration which influenced how they sustained momentum and

 © ISLS

supported one another. By planning, measuring, troubleshooting, and hypothesizing together, students engaged in authentic scientific practices.

Thematic analysis of students' SSRL observed in video data

The first and second authors used an interaction analysis (Jordan & Henderson, 1995) to examine both theoretically grounded and contextually emergent aspects of SSRL in the video data. The interpretations were anchored in what was directly observable in the video record (e.g., talk, gestures, use of tools) and had room for unexpected patterns of interaction that emerged beyond the initial analytic frame. We connected theoretical constructs of SSRL with the situated practices visible on video. Our initial analytic frame drew on established SSRL literature, and our focus was structured around cognitive and metacognitive regulation (Hadwin et al., 2018; Isohätälä et al., 2017; Panadero & Järvelä, 2015). Building on this foundation, we adapted an exploratory stance to refine and expand categories in response to recurring interactional features observed in the data.

For eight hours of interactional analysis sessions, we watched the videos together, paused recordings at key moments, and took individual memos. Pausing and replaying the recordings allowed us to observe verbal and nonverbal behaviors carefully, trace actions in sequence, and challenge assumptions through repeated review. At the conclusion of each hour-long session, the first author synthesized both sets of memos into an integrated record to identify recurring practices observed in the clips. The synthesized memos structured the findings into three parts: (1) key instances categorized into the cognitive and metacognitive dimensions of SSRL; (2) key practices with direct quotes and interpretations; (3) emergent overarching themes from researcher discussion and memo comparison. These synthesized memos served as a living document to iteratively test, refine, and revise category definitions across sessions. Joint viewing and memo exchange helped us surface and question different interpretations, counteract confirmation bias and reveal unnoticed aspects of interaction. By documenting individual perspectives, comparing observations, and negotiating interpretations, we worked to surface assumptions and strengthen the trustworthiness of our claims.

Findings

The analysis across three groups revealed that SSRL emerged most prominently during moments of uncertainty, instances where students were unsure how to generate, interpret, or verify knowledge in the midst of fieldwork. These moments, triggered by unclear task goals, unexpected data patterns, or technological malfunctions prompted groups to engage in reciprocal cognitive regulation to sustain inquiry. Key indicators included questioning, help-seeking, reframing of information, and negotiated explanations with monitoring distributed across members. For instance, three 11th-grade female students collecting indoor air-quality data across four school sites encountered a sensor that failed to update its location, creating uncertainty about whether their measurements were registering correctly:

> Student 1: "Wait, refresh data."
>
> Student 2: "Doesn't do anything."
>
> Student 1: "I mean, our location definitely did change, but please, no, it did not [on the sensor.] I mean, are we doing something wrong?
>
> Student 2: "Oh?! Oh my God."
>
> Student 3: "Did it change?"
>
> Student 1: "No, the location is definitely changing, but this is,"
>
> Student 3: "Should we exit and go back?"
>
> Student 1: "Strange, like, I didn't think it was strange for the first two, but we're in a whole different area right now."
>
> Student 2: "This battery is out"
>
> Students 1, 3: "What? Nooo."
>
> Student 1: "Maybe that's why it wasn't changing."

This exchange moved from status-checking to error recognition to shared sensemaking. The students could not determine whether the problem was procedural or technical, and that uncertainty drove collective hypothesis generation. A hypothesis on low battery emerged and was linked to the anomaly which converted confusion into an actionable explanation. This shows cognitive SSRL as a cycle of identifying a problem, generating a hypothesis, and aligning explanations. The mobile sensor was the source of uncertainty that catalyzed the group's joint reasoning.

Planning and strategizing under constraint represented another form of cognitive SSRL. For example, when three 12th-grade students arrived to find their intended data collection site locked, they revised their plan collectively:

> Student 1: "It's a padlock."
>
> Student 3: Okay, we need to find [a new location] real quick. Actually, we should choose any spot we want.
>
> Student 2: "Yeah, you got any good spots? [We can] go to the bathroom, like everybody else."
>
> Student 3: "We can go to the lunch room. We can go to the stairway. Stairway [is] usually a pretty interesting one that's, like, really humid..."
>
> Student 2: "Yeah, are you talking about [the one near the even side on the corner]? Yeah, I forgot where it is, specifically in the basement."
>
> Student 1: "Um, actually, why don't we go to the kitchen? … I mean we're here to test the humidity [that affects air quality] … The kitchen might be one of the better places because there's so much machinery and stuff running."
>
> Student 2: "Yeah, right."

Students pooled local knowledge, evaluated candidate sites, and selected the kitchen by connecting its machinery use to their investigation goals. This illustrates cognitive SSRL as goal-directed planning: the group self-organized around a shared problem without teacher intervention or external support.

Across both episodes, uncertainty was converted into productive joint reasoning. Students distributed monitoring, externalized doubts, generated and tested hypotheses, and aligned on next steps. Their practices were core features of cognitive SSRL that kept inquiry coherent despite real-world constraints and positioned uncertainty as a resource for collaborative scientific reasoning.

Discussion and conclusion

The findings suggest that SSRL in the authentic STEM learning environment was driven by transforming uncertainty into collective cognitive problems. The most generative SSRL episodes arose during breakdowns including device failures, inconsistent data, blocked sites. These tensions triggered cycles of monitoring, troubleshooting, and sense-making. As a constitutive feature of scientific practice and a pedagogical resource for fostering student engagement, uncertainty facilitated student inquiry (Chen et al., 2024; Manz, 2018).

These findings extend scholarship on SSRL and productive uncertainty. Chen et al. (2024) argue that epistemic uncertainty drives students toward hypothesis generation and evidence construction. The current study shows this process operating at the group level with monitoring distributed and explanations negotiated collectively through transactive exchanges (Hadwin et al., 2011; Isohätälä et al., 2017; Järvenoja et al., 2020) and reciprocal participation and contribution (Isohätälä et al., 2017; Törmänen et al., 2022). Similarly, Manz et al. (2025) argue that students who engage productively with uncertainty develop more agency in scientific practice. The findings show that this collective agency can emerge organically without external intervention. Schnaubert's (2023) insight that voicing uncertainty triggers supportive regulation is also confirmed. In both episodes, publicizing the problem initiated the group's regulatory cycle.

For CSCL, this study offers design principles for place-based, mobile-technology-mediated STEM learning environments as sites for authentic scientific inquiry. We show how cognitive moves evolve within short episodes, extending SSRL beyond classroom talk to the engagement with environments and tools in collaborative groups (Jeong et al., 2019). First, structured autonomy supports a shift from teacher-reliant troubleshooting to peer-led regulation. Second, authentic scientific inquiry turns technology into both mediator and catalyst where breakdowns become opportunities for learning. Lastly, place-based resources broaden CSCL's focus from technology use to spatial orchestration. The findings yield actionable principles including pairing minimal roles with real choice over sites and procedures, making uncertainty legible and consequential, and cultivating agentive norms that celebrate micro-wins, clarifying how to turn uncertainty into a resource for collaborative learning.

The study has limitations that future work should address. The sample is small and drawn from a specific local urban context, limiting generalizability. The analysis focuses on cognitive SSRL; the behavioral and socio-emotional dimensions co-present in these episodes warrant dedicated investigation. Future research should examine the optimal level and type of epistemic uncertainty for supporting SSRL across different group compositions and inquiry phases, and whether design interventions are most effective when targeting individual metacognitive awareness or the group's collective regulatory capacity.

References

Chen, Y., Jordan, M., Park, J., & Starrett, E. (2024). Navigating student uncertainty for productive struggle: Establishing the importance for and distinguishing types, sources, and desirability of scientific uncertainties. *Science Education*, sce.21864.

Dragnić-Cindrić, D., Greene, J., & Anderson, J. (2020). The role of uncertainty in social regulation of learning. In Gresalfi, M. and Horn, I. S. (Eds.), *The Interdisciplinarity of the Learning Sciences, 14th International Conference of the Learning Sciences (ICLS) 2020, Volume 2* (pp. 759-760). Nashville, Tennessee: International Society of the Learning Sciences.

Hadwin, A., Järvelä, S., & Miller, M. (2018). Self-regulation, co-regulation, and shared regulation in collaborative learning environments. In D. H. Schunk & J. A. Greene (Eds.), Handbook of self-regulation of learning and performance (2nd ed., pp. 83-106). Routledge.

Institute of Education Sciences (IES) & National Science Foundation (NSF). (2013). *Common Guidelines for Education Research and Development.*

Isohätälä, J., Järvenoja, H., & Järvelä, S. (2017). Socially shared regulation of learning and participation in social interaction in collaborative learning. *International Journal of Educational Research, 81*, 11-24.

Järvenoja, H., Järvelä, S., & Malmberg, J. (2020). Supporting groups' emotion and motivation regulation during collaborative learning. *Learning and Instruction, 70*, 101090.

Jeong, H., Hmelo-Silver, C. E., & Jo, K. (2019). Ten years of Computer-Supported Collaborative Learning: A meta-analysis of CSCL in STEM education during 2005–2014. *Educational Research Review, 28*, 100284.

Jordan, B. & Henderson, A. (1995) Interaction analysis: Foundations and practice, *Journal of the Learning Sciences, 4*(1), 39-103.

Kim, J., Detrick, R., Yu, S., Song, Y., Bol, L., & Li, N. (2025). Socially shared regulation of learning and artificial intelligence: Opportunities to support socially shared regulation. *Education and Information Technologies, 30*(9), 11483–11521.

Manz, E. (2018). Designing for and analyzing productive uncertainty in science investigations. In Kay, J. and Luckin, R. (Eds.) Rethinking Learning in the Digital Age: Making the Learning Sciences Count, *13th International Conference of the Learning Sciences (ICLS) 2018, Volume 1*. London, UK: International Society of the Learning Sciences.

Manz, E., Ward, A., & Wells, A. (2025). Connecting design and experience: Tracing uncertainty through classroom science investigations. *Science Education*, sce.70015.

Michalsky, T., & Cohen, A. (2021). Prompting socially shared regulation of learning and creativity in solving STEM problems. *Frontiers in Psychology, 12(722535)*, 1-12.

Panadero, E., & Järvelä, S. (2015). Socially shared regulation of learning: A review. *European Psychologist, 20*(3), 190-203.

Schnaubert, L. (2023). Uncertainty awareness in digital and social learning environments. In Slotta, J. D. & Charles, E. S. (Eds.). *General Proceedings of the 3rd Annual Meeting of the International Society of the Learning Sciences 2023*. Montreal, Canada: International Society of the Learning Sciences.

Sharma, K., Nguyen, A., & Hong, Y. (2024). Self-regulation and shared regulation in collaborative learning in adaptive digital learning environments: A systematic review of empirical studies. *British Journal of Educational Technology, 55*(4), 1398–1436.

Sulla, F., Monacis, D., & Limone, P. (2023). A systematic review of the role of teachers' support in promoting socially shared regulatory strategies for learning. *Frontiers in Psychology*, 14, 1208012.

Törmänen, T., Järvenoja, H., & Järvelä, S. (2023). Affective states and regulation of learning during socio-emotional interactions in secondary school collaborative groups. *British Journal of Educational Psychology, 93*(1), 89-108.

Yoon, S. A., Miller, K. M., Richman, T., Noushad, N., Hageman, G., Liu, Y., Zhang, W., & Cottone, A. M. (2025). Developing teachers' disciplinary knowledge for high school STEM integration: A review of a decade of educational research. *Review of Educational Research, 95*(6), 1337–1378.

Yoon, S. A., Shim, J., Miller, K., Cottone, A. M., Noushad, N. F., Yoo, J., Gonzalez, M. V., Urbanowicz, R., & Himes, B. E. (2022). Professional development for STEM integration: Analyzing bioinformatics teaching through considerations of adaptive expertise and cognitive load. In Superfine, A.C., Goldman, S. R., & Ko, M. M. (Eds.), *Teacher learning in changing contexts: Perspectives from the learning sciences* (pp. 69–90). Routledge.

International Society of the Learning Sciences

MIRACLE: Multi-Agent Intelligent Regulation to Advance Collaborative Learning Environment

Shuang Li, Haiyang Xin, Yimeng Sun, Qiannan Niu
lishuang@cocorobo.cc, Tony@cocorobo.cc, sunyimeng@cocorobo.cc, niuqiannan@cocorobo.cc,
COCOROBO Limited

Lingyun Huang, The Education University of Hong Kong, lingyunhuang@eduhk.hk
Gaowei Chen, The University of Hong Kong, gwchen@hku.hk
Yibing Zhang, Nanjing Normal University, zyc@njnu.edu.cn
Ching Sing Chai, The Chinese University of Hong Kong, CSChai@cuhk.edu.hk

Abstract: Effective collaboration requires Socially Shared Regulation (SSRL), but students often lack these skills. This study introduces the MIRACLE (Multi-Agent Intelligent Regulation to Advance Collaborative Learning Environment) system, which supports SSRL by orchestrating metacognitive regulation and proactively providing emotional and motivational support. We conducted a quasi-experimental study with 90 fifth-grade students. The experimental group (n=42) used a collaborative platform CocoNote equipped with MIRACLE, while the control group (n=48) used the same platform with a general GPT assistant. Quantitative results show the MIRACLE group achieved significant gains across SSRL phases (Planning, Monitoring, Reflection) and produced higher-quality collaborative artifacts compared to the control group. Qualitative findings indicate students perceived MIRACLE as an effective facilitator for cognitive, regulatory, and emotional support. This study demonstrates that specialized, orchestrated AI systems are more effective than generic AI in enhancing SSRL.

Introduction

Collaboration has been widely recognized as a powerful pedagogical approach that enhances students' academic performance, knowledge construction, and problem-solving skills (Miller & Hadwin, 2024). However, effective collaboration is far more complex than individual learning and is often hindered by challenges such as unequal participation, dominance of certain members, and difficulties in reaching consensus (Zheng et al., 2023). A major underlying reason for these challenges is the insufficient development of Socially Shared Regulation of Learning (SSRL)—the collective process through which group members plan, monitor, and regulate their joint learning activities (Järvenoja et al., 2020). While SSRL is essential for productive collaboration, many learners lack the regulatory competence to coordinate their efforts effectively, thereby requiring structured guidance and external support (Järvelä et al., 2013).

Earlier research on technology-enhanced collaborative learning provided such structured support through collaboration scripts and group awareness tools (Järvelä et al., 2015; Miller & Hadwin, 2015). With the rise of Generative Artificial Intelligence (GenAI), new opportunities have emerged to offer more flexible and context-aware support. Leveraging GenAI's strengths in language understanding and dynamic feedback, researchers have begun developing conversational agents capable of delivering personalized and "in-the-moment" guidance (Kasneci et al., 2023). Edwards (2024) introduced a proactive speech agent that prompts students to reflect and regulate during group discussions. While the proactive agent demonstrates potential, it provides limited support types and may not respond adequately to students' diverse and evolving needs. Recent multi-agent systems (Xin et al., 2025) attempt to expand functionality through the integration of multiple agents (e.g., proactive and reactive agents). However, this multi-agent configuration introduces a new challenge: uncoordinated interactions with multiple agents can increase students' cognitive load and lead to fragmented or inconsistent learning experiences. Thus, there remains a pressing need for intelligently orchestrated agent systems that can dynamically coordinate different regulatory functions in a structured and seamless manner.

To address this gap, this study presents MIRACLE (Multi-Agent Intelligent Regulation to Advance Collaborative Learning Environments)—an orchestrated multi-agent system designed to enhance SSRL by combining metacognitive regulation with proactive emotional and motivational support. Implemented within a collaborative learning platform CocoNote, MIRACLE aims to provide students with adaptive and structured assistance. This study investigates:

RQ1: Does the MIRACLE system enhance students' SSRL abilities and collaborative learning outcomes?

RQ2: How do students perceive the MIRACLE system and its influence on their collaborative learning?

 © ISLS

Theoretical foundations

Socially shared regulation of learning

Derived from self-regulated learning (SRL) theory, SSRL is the process through which group members consciously regulate their collective cognitive, behavioral, motivational, and emotional states. This management occurs through an iterative cycle of planning, monitoring, evaluating, and regulating phases (Hadwin et al., 2017). The MIRACLE system is engineered to foster and maintain this SSRL process by delivering regulatory support for the planning, monitoring, and reflection stages, alongside cognitive assistance during task execution.

Trigger Regulation Framework

The Trigger Regulation Framework (Järvelä & Hadwin, 2024) identifies triggering events as critical moments for supporting learners' collaboration. These triggering events refer to circumstances or behavioral patterns that impede learning progress, such as diminished participation or adverse emotional states, which present valuable opportunities for facilitating metacognitive development (Edwards et al., 2024). Drawing upon this framework, the MIRACLE system's proactive agents were designed to promote collective metacognitive awareness and participation through the provision of metacognitive, socio-emotional, and motivational scaffolding.

The collaborative learning platform CocoNote

The collaborative learning platform CocoNote integrates a hypermedia whiteboard, a group chatroom, and the MIRACLE system to facilitate SSRL. The whiteboard allows learners to co-construct knowledge by creating and connecting multimodal notes (text, images, and videos), while the chatroom enables synchronous communication for task deliberation and coordination.

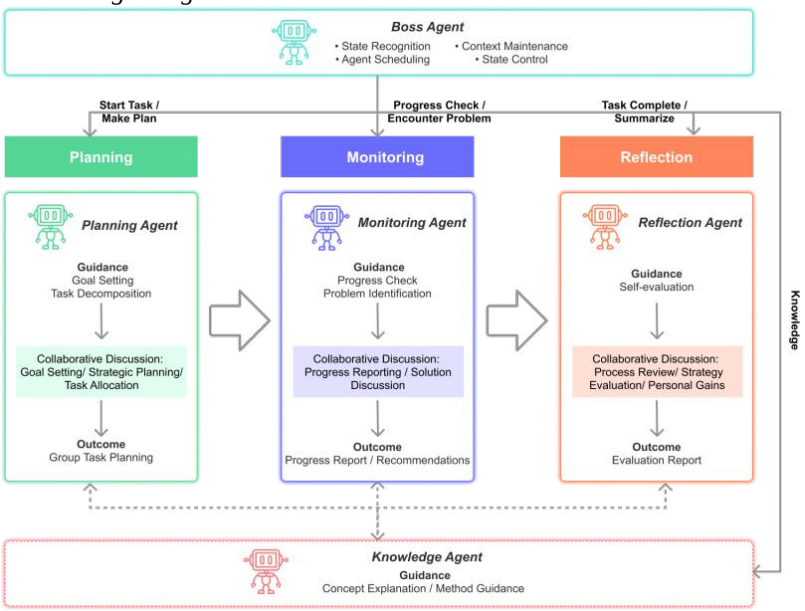
The MIRACLE agent system design

The MIRACLE system (implemented within the platform) is powered by GPT-5-Nano and incorporates both reactive agents and one proactive agent to support the SSRL process.

Reactive Agents. The reactive agents employ a hierarchical agent architecture built using the LangGraph framework (LangChain, 2024). At its core is a Boss Agent that orchestrates all student-agent interactions by intelligently routing student requests to four specialized agents: the Planning Agent, Monitoring Agent, Reflection Agent, and Knowledge Agent (see Figure 1). When students mention the Boss Agent in the chatroom, it analyzes their messages to determine whether they need knowledge support or metacognitive guidance, then directs them to the appropriate specialized agent. All agents access shared real-time data from both the chatroom and whiteboard, enabling contextually grounded responses. This design allows students to focus on collaborative tasks rather than navigating between different agents.

Figure 1

The Intelligent Agent Orchestration Architecture



 © ISLS

Proactive Agent. Recognizing the critical role of emotional and motivational factors in sustaining collaborative engagement, particularly among younger learners, the MIRACLE system also features a dedicated Lightbulb Agent that operates independently from the metacognitive support agents. The Lightbulb monitors both chatroom conversations and whiteboard activities to detect triggering events that signal emotional or motivational challenges, such as extended periods of individual inactivity, expressions of confusion or frustration, declining participation rates, or absence of progress in task execution. Unlike reactive agents that respond to explicit student requests, the Lightbulb operates proactively, initiating interventions when triggering events are detected. Upon identifying a trigger, the agent provides feedback via an interface and notifies learners by flashing a "lightbulb" icon on the top-right corner of the whiteboard. Once learners view the prompt, the lightbulb stops flashing until the next trigger.

Research design and methods

This quasi-experimental study involved 90 fifth-grade students (aged 10–11) from two parallel classes in an elementary school from a coastal city with well-developed educational infrastructure in Southern China. The experimental group (n = 42) used the collaborative platform equipped with the MIRACLE system, while the control group (n = 48) used the same platform—with identical whiteboard and chatroom features—but interacted with a general GPT assistant (also powered by GPT-5-Nano) instead of MIRACLE. The GPT assistant received a general-purpose system prompt without specialized pedagogical instructions and operated in a purely reactive mode, responding only when students mentioned it in the chatroom; it had no proactive triggering mechanism or orchestrated agent routing. Students worked in groups of 4–6 to complete a 120-minute design-thinking project in the computer lab, which required them to plan and present an innovative gift design, including its concept, appearance, materials, and production process. Multiple data sources were collected to ensure validity. The Socially Shared Regulated Learning Questionnaire (Yang et al., 2025) was administered before and after the intervention to assess students' SSRL abilities. Group artifacts were collected and evaluated for quality, and six groups (three from each condition) participated in post-study interviews to provide qualitative insights into their collaborative experiences.

Data treatment

To examine intervention effects on SSRL abilities (RQ1), we performed 2×2 repeated-measures ANOVAs on four SSRL dimensions (Yang et al., 2025), with paired t-tests for within-group changes. Two raters assessed artifacts (N=18) using a three-dimensional rubric (ICC=.89). Semi-structured interviews with six groups (three experimental, three control) were analyzed using reflexive thematic analysis (Cohen's κ =.80).

Findings

RQ1. Does the MIRACLE system enhance students' SSRL abilities and collaborative learning outcomes?

Repeated measures ANOVA revealed significant Time × Group interactions for Planning and Monitoring (see Table 1). The experimental group improved significantly on Planning, Monitoring, and Reflection, while the control group showed no changes. Emotion Regulation showed no significant effects.

Table 1

Pre-Post Changes and Intervention Effects on SSRL Dimensions

Dimension	Group	Pre-test	Post-test	Within-Group Change		Time × Group Interaction		
		M ± SD	M ± SD	t	p	F	p	η^2p
Planning	Experimental	26.43±5.92	28.38±4.38	-2.07	.045*	7.44	.008**	.078
	Control	25.75±5.41	24.19±6.26	1.78	.082			
Monitoring	Experimental	24.48±6.81	28.43±4.79	-3.09	.004**	7.15	.009**	.075
	Control	25.46±6.30	25.15±7.55	0.32	.753			
Reflection	Experimental	24.79±5.56	27.10±4.80	-2.28	.028*	3.21	.077†	.035
	Control	23.96±5.15	24.06±5.96	-0.14	.888			
Emotion	Experimental	46.14±8.62	47.45±6.54	-0.92	.365	1.45	.231	.016
	Control	45.08±7.03	44.27±8.87	0.76	.452			

 © ISLS

Mann-Whitney U tests were conducted to compare artifact quality between experimental and control groups, as the data violated normality assumptions (Shapiro-Wilk test, $p < .05$). Results revealed that experimental groups (Mdn = 7.50, M = 7.61, SD = 0.99) produced significantly higher-quality artifacts than control groups (Mdn = 5.50, M = 4.78, SD = 1.48), $z = -3.27$, $p < .001$, $r = .77$, representing a large effect size. Dimensional analyses showed that experimental groups outperformed control groups across all three aspects.

RQ2.How do students perceive the MIRACLE system and its influence on their collaborative learning?

Semi-structured interviews with six groups revealed how students perceived the multi-agent system's influence on their collaborative learning experiences.

Boss agent orchestration: Cognitive support and process regulation

Experimental students perceived the Boss Agent as a holistic facilitator that streamlined both their learning process and their group collaboration. Its value spanned the entire SSRL cycle, manifesting in distinct roles at each stage. For initiating the task, it served as both an ideation partner for planning and an on-demand expert for knowledge support. This support then transitioned into a supervisory role during task execution. For monitoring, the agent acted as an objective third-party supervisor, prompting adjustments when it "said our progress wasn't particularly good... we took the suggestions and things got much better" (Student C). For reflection, it functioned as an administrator, providing comprehensive summaries that students "felt... was very thorough" (Student D).

Crucially, students reported that this automated, objective support was the key mechanism for unifying group decisions and reducing conflicts. By acting as a central coordinator for the team's SSRL process, the Boss Agent made the collaboration more orderly, experienced by students as seamless assistance that "really sped up our task progress" (Student C).

Lightbulb agent: Emotional support and member activation

Students generally perceived the Lightbulb Agent as providing valuable emotional support rather than intrusion. Students appreciated its personalized nature, noting that its suggestions were highly relevant to their specific contributions: "The Lightbulb knows what I have made on the white board, its suggestion is relevant to my actual situation" (Student E).

Most notably, the Lightbulb's visible interventions provided crucial socio-emotional scaffolding, fostering a sense of unity and activating disengaged members. Its prompts were seen as warm encouragement that helped students engage in teamwork: "It told me no need to be anxious and just take my time" (Student F), and motivated reluctant participants to rejoin the collaborative effort: "At first, some people in our group weren't doing anything. But when they saw the Lightbulb... they started cooperating with us" (Student G).

Contrasting experiences and overall impact

The positive perceptions of the MIRACLE system stand in stark contrast to the experiences of the control group. Experimental students unanimously affirmed that their collaboration improved: "Yes, it got better... things became more organized" (Student H).

In contrast, control group students reported significant coordination failures: "It didn't go well... mainly because we weren't united, and the group leader we chose didn't have management ability" (Student I). Notably, their primary desire was for basic oversight mechanisms—"The platform should remind or report students who go off-topic" (Student J). This divergence suggests that MIRACLE satisfied foundational regulatory needs, allowing groups to focus on higher-level collaboration.

Discussion, limitations and future directions

The experimental group's consistent gains in Planning, Monitoring, and Reflection—the three core phases of SSRL—demonstrate that MIRACLE's pedagogically intelligent and orchestrated agents effectively scaffolded the entire regulatory process. In contrast, the control group, despite using a general GPT assistant within the same collaborative environment, showed little progress, reinforcing the necessity of structured, pedagogical integration over generic AI assistance for fostering meaningful regulation in collaboration. The lack of significant quantitative gains in Emotion Regulation, despite qualitative evidence of emotional support, likely reflects that emotional regulation competencies require longer internalization periods than the 120-minute intervention provided. The study's small, age-specific sample and short intervention duration limit generalizability and long-term inference. Future research should employ larger samples, longer interventions, and multimodal data to deepen understanding of pedagogical AI's impact on collaborative regulation.

References

Edwards, J., Nguyen, A., Lämsä, J., Sobocinski, M., Whitehead, R., Dang, B., Roberts, A., & Järvelä, S. (2025). Human-AI collaboration: Designing artificial agents to facilitate socially shared regulation among learners. *British Journal of Educational Technology*, *56*(2), 712–733.

Edwards, J., Nguyen, A., Sobocinski, M., Lämsä, J., De Araujo, A., Dang, B., ... & Jarvela, S. (2024, July). MAI-A Proactive Speech Agent for Metacognitive Mediation in Collaborative Learning. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces* (pp. 1-5).

Hadwin, A., Järvelä, S., & Miller, M. (2017). Self-Regulation, Co-Regulation, and Shared Regulation in Collaborative Learning Environments. In D. H. Schunk & J. A. Greene (Eds.), *Handbook of Self-Regulation of Learning and Performance* (2nd ed., pp. 83–106). Routledge.

Järvelä, S., & Hadwin, A. (2024). Triggers for self-regulated learning: A conceptual framework for advancing multimodal research about SRL. *Learning and Individual Differences*, *115*, 102526.

Järvelä, S., Järvenoja, H., Malmberg, J., & Hadwin, A. F. (2013). Exploring socially shared regulation in the context of collaboration. *Journal of Cognitive Education and Psychology*, *12*(3), 267-286.

Järvelä, S., Kirschner, P. A., Panadero, E., Malmberg, J., Phielix, C., Jaspers, J., Koivuniemi, M., & Järvenoja, H. (2015). Enhancing socially shared regulation in collaborative learning groups: Designing for CSCL regulation tools. *Educational Technology Research and Development*, *63*(1), 125–142.

Järvenoja, H., Järvelä, S., & Malmberg, J. (2020). Supporting groups' emotion and motivation regulation during collaborative learning. *Learning and Instruction*, 70, 101090.

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., . . . Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, 102274. https://doi.org/10.1016/j.lindif.2023.102274

LangChain. (2024). *LangGraph*. https://www.langchain.com/langgraph

Miller, M., & Hadwin, A. (2015). Scripting and awareness tools for regulating collaborative learning: Changing the landscape of support in CSCL. *Computers in Human Behavior*, 52, 573-588.

Miller, M., & Hadwin, A. F. (2024). Comparing the effectiveness of CSCL scripts for shared task perceptions in socially shared regulation of collaborative learning. *International Journal of Computer-Supported Collaborative Learning*, 1-24.

Xin, H., Li, S., Huang, L., Yip, V. W., & Liu, J. (2025). CocoNote: Agents-aided Collaborative Learning Environment Enhances Socially Shared Regulation. In 18th International Conference on Computer-Supported Collaborative Learning (CSCL) 2025-CSCL Proceedings.

Yang, Y., He, Z., Wei, Y., & Tang, E. (2025). Development and validation of the socially shared regulated learning questionnaire: insights from a second-order confirmatory factor analysis. *Frontiers in Psychology*, 16, 1635325

Zheng, L., Long, M., Chen, B., & Fan, Y. (2023). Promoting knowledge elaboration, socially shared regulation, and group performance in collaborative learning: An automated assessment and feedback approach based on knowledge graphs. *International Journal of Educational Technology in Higher Education*, *20*(1), 46.

International Society of the Learning Sciences

Unveiling Role Dynamics in AR-Based Collaborative Inquiry

Xinyue Jiao, New York University, xj2320@nyu.edu
Zifeng Liu, University of Florida, liuzifeng@ufl.edu

Abstract: Augmented Reality (AR) has emerged as a transformative tool in education, offering authentic and interactive environments for collaborative inquiry. However, collaborative interaction among group members in AR-based inquiry, particularly how roles are enacted and exchanged, remains underexplored. This study investigates role interaction patterns in AR-supported collaborative inquiry and analyzes how role dynamics between group members influence learning outcomes. Eighty students working in pairs engaged in AR-based scientific inquiry tasks. Video recordings were coded and analyzed. Cluster analysis shows that three distinct patterns emerged: Balanced, Dominant, and Passive. The Balanced pattern featured reciprocal *Question–Respond–Experiment–Share* behavioral sequences; Dominant dyads exhibited leader–follower dynamics, while Passive dyads demonstrated minimal engagement and weak regulation. Quantitative results showed that active participation, regardless of balance or dominance, was associated with higher learning outcomes. These findings advance understanding of how role dynamics unfold in immersive learning environments, offering insights for promoting effective collaboration.

Introduction and related work

Augmented Reality (AR) has emerged as a promising tool for collaborative inquiry learning by embedding scientific exploration in interactive and embodied environments (Connor et al., 2024; Dunleavy et al., 2009; Li et al., 2025; Mansour et al., 2024). Prior studies have integrated AR into scientific inquiry-based learning, demonstrating that AR supports students' conceptual understanding, motivation, and engagement by allowing them to visualize and manipulate abstract phenomena (Silva et al., 2023; Upadhyay et al., 2024; Zhang et al., 2025). However, while prior studies have primarily focused on learning outcomes, limited research has explored how collaborative inquiry processes unfold in AR settings (Anna & Spyros, 2025; Buchner & Kerres, 2023). Although recent research has explored overall collaboration process in AR-supported inquiry, for example, revealing distinct modes of coordination and participation (Connor et al., 2024) or how AR affects collaboration learning (Radu & Schneider, 2022), less attention has been paid to the micro-level interaction processes within groups, particularly how roles emerge, are exchanged, and interact with one another during AR inquiry, and how these dynamics relate to learning outcomes.

The emergence and transformation of roles are critical because they reflect students' participation structures and shape the flow of collaboration (Barron, 2003; Dowell et al., 2020; Jiao et al., 2022). While role formation and transitions have been extensively examined in face-to-face and computer-supported collaborative learning (CSCL) contexts (Fischer et al., 2013; Saqr et al., 2024; Srba & Bielikova, 2014), AR introduces unique coordination challenges, such as sharing devices, manipulating digital objects, and managing visual attention between physical and virtual spaces, which can significantly affect role distribution and participation balance (Radu & Schneider, 2022). These multimodal demands may increase cognitive load and reshape how roles emerge, are exchanged, and interact during inquiry. Understanding these role dynamics is essential for designing adaptive AR learning environments that foster productive collaboration.

To address these gaps, this study aims to understand how role dynamics unfold in AR learning environments. Specifically, the research questions are as follows: (1) RQ1: What types of role interaction patterns emerge among students during AR-based collaborative inquiry, and what characterizes each? (2) RQ2: How are these role interaction patterns related to students' learning outcomes? By identifying how different role interaction patterns emerge and how they relate to learning outcomes, educators and designers can develop scaffolds that promote balanced participation and deeper inquiry learning.

Methods

Participants and AR collaborative learning environment

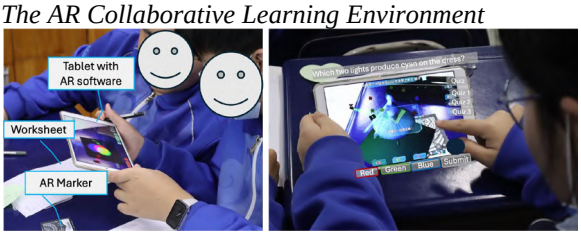
This study involved 80 seventh-grade students (37 boys, 43 girls) from two middle school classes. The intervention lasted two weeks. In the first week, students completed a 15-minute pretest designed to assess their prior knowledge of scientific concepts related to light. Subsequently, an instructor delivered a brief introduction to AR technology. During the second week, students worked in pairs (40 dyads) to complete AR inquiry tasks

 © ISLS

using a tablet. The inquiry topic is the mixing rules of primary colors in physics. After the session, students completed a post-test. This study developed an AR application, The Light and Color, to support interactive collaborative learning of the principles of the three primary colors of light (see Figure 1). The AR application created experimental scenarios in which students could explore light color mixing by manipulating both on-screen controls and physical markers. Students then worked in pairs to conduct the AR-based experiments collaboratively and record their scientific observations in a worksheet.

Figure 1

The AR Collaborative Learning Environment



Data collection and analysis

For RQ1, the collaborative processes of 40 groups were captured using video recordings. A rating scheme was constructed based on Storch's (2002) dyadic interaction model, which centers on the two dimensions of equality and mutuality. Equality refers to the balance of participation between partners and was further operationalized through indicators such as role distribution, role switching, and contribution balance. Mutuality reflects the degree of connectedness and responsiveness in partners' contributions and was operationalized through indicators including reciprocity, group coordination, and group cohesion. All indicators were rated on a 5-point Likert scale. To establish reliability, two researchers independently coded 30% of the videos, achieving strong agreement (Cohen's $\kappa = 0.72$). After that, one researcher coded the remaining data. Subsequently, K-means clustering was applied to the equality–mutuality ratings to identify distinct role interaction patterns. To further explore the characteristics of each pattern, four representative dyads from each cluster were selected for in-depth qualitative analysis, focusing on role distribution, role exchange, and interaction frequency features. To gain a deeper understanding of the interactive behaviors within each dyad, group interactions were coded by two researchers using the Collaborative Inquiry Interaction Behavior Coding Scheme, which included the following categories: *Question*, *Respond*, *Share*, *Experiment*, *Analyze*, *Negotiate*, *Consensus*, *Organization*, and *Off-task*, adapted from Hou & Wu (2011)'s social interaction framework. Inter-rater reliability was high (Cohen's $\kappa = 0.77$). Any discrepancies were examined collaboratively and resolved through consensus. During coding, the two roles in each dyad were differentiated using role-specific labels (e.g., A_Ques, B_Resp), where A and B represent each participant. Next, Sequential Pattern Mining was conducted to examine the sequential transitions of behaviors between members. A Contiguous N-gram Mining approach was used to identify frequent behavioral sequences, with the N-gram length set between 2 and 5 to capture both short and extended interaction cycles.

For RQ2, students' collaborative inquiry outcomes were assessed using scientific knowledge tests, including pre-test and post-test. Each assessment was scored out of 100 points. Scoring reliability was ensured by two researchers, with a Kappa coefficient of 0.95. ANOVA test was used to compare intergroup differences in knowledge test scores across the identified patterns.

Results

Role Interaction patterns in AR collaborative inquiry

To address RQ1, we applied K-means clustering to categorize the role interaction characteristics of 40 groups. The elbow method was used to determine the optimal number of clusters, and the solution with $K = 3$ provided the best explanatory power. Cluster 1 ($n = 14$) was characterized by high scores on both equality and mutuality, indicating balanced participation and reciprocal interaction between partners. This cluster was therefore labeled the Balanced pattern. Cluster 2 ($n = 14$) showed moderate equality and mutuality. With limited role switching and uneven contribution balance, this cluster reflected a leader-driven interaction structure and was labeled the Dominant pattern. Cluster 3 ($n = 12$) scored low on both equality and mutuality, suggesting limited participation and weak collaboration among members, and was labeled the Passive pattern.

To further examine the specific characteristics of these role interaction patterns, four representative groups from each cluster were selected for further analysis. The types of roles enacted during collaboration were first identified, followed by an examination of role distribution and exchange features across clusters. Additionally,

 © ISLS

we analyzed the sequential dynamics of role exchange for one representative group from each cluster, as shown in Figure 2. Finally, the interaction frequencies between roles within each dyad were recorded (See Table 1).

Table 1

Role Distribution and Interaction of Three Types

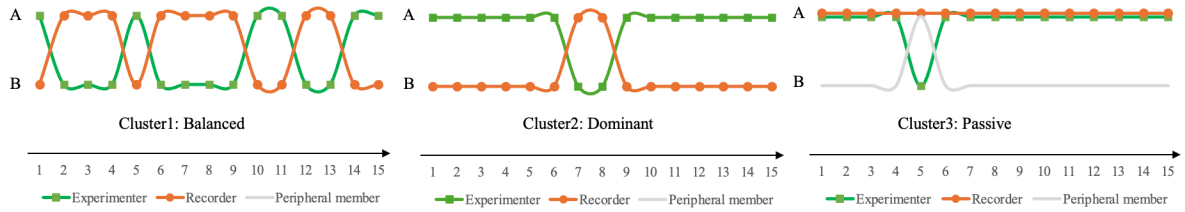
Patterns	Group No.	Role distribution		Role exchange	Role interaction Frequency (f)	
		Member A	Member B	Frequency	A → B	B → A
Balanced	9	Experimenter+ recorder	Experimenter+recorder	6	29	30
	13	Experimenter+ recorder	Experimenter+recorder	5	22	18
Dominant	1	Leader+ Experimenter	Recorder	2	25	15
	8	Experimenter	Recorder + Observer	1	18	11
	6	Recorder	Experimenter	1	12	20
Passive	12	Peripheral Member	Experimenter+Observer	1	5	16
	4	Experimenter	Peripheral Member	2	14	7

Note: Not a full list of all groups

Based on Table 1 and Figure 2, we can find that (1) Balanced Pattern. Dyads in this cluster showed reciprocal and coordinated role engagement. Both members alternated between being the Experimenter and Recorder, demonstrating frequent role exchange (f = 5–6) and nearly symmetrical interaction frequencies between group members (A → B and B → A). This pattern reflected mutual contribution and shared regulation, indicating that both partners actively participated in experimentation and documentation. (2) Dominant Pattern. Groups in this cluster exhibited a fixed role characteristic, where one member (A) consistently acted as the Leader or Experimenter, while the partner (B) primarily took the Recorder or Observer role. Role exchange occurred infrequently (f = 1–2), and interaction flows were asymmetric, with more messages or actions initiated by the dominant member, for example, the frequency of A → B (f) is more than B → A (f) in group 1. From the observation, we can find that role interactions in this pattern were predominantly unidirectional, with the Experimenter operating the AR system and reporting findings to the Recorder, showing an imbalanced structure and static task division. (3) Passive Pattern. This cluster featured low engagement and minimal coordination. Interaction and communication between members were minimal. One member consistently performed both the Experimenter and Recorder roles, while the other remained a passive observer or peripheral participant, showing limited initiative or response, which is a highly unbalanced and ineffective collaborative structure.

Figure 2

The Role Exchange Dynamics Over Time



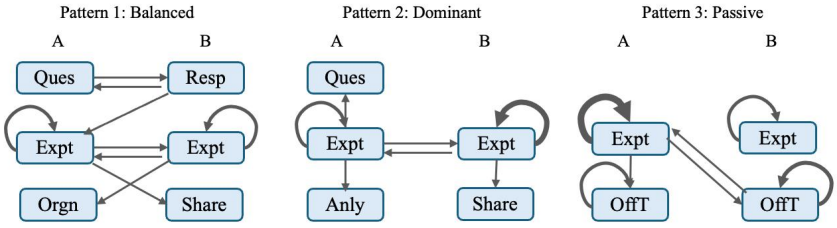
To further understand the interactive behaviors between group members, we conducted sequential pattern mining to examine the temporal structures of roles' actions. This analysis revealed the most frequent behavioral transitions between partners. For example, the transition *A_Question → B_Respond (f = 24)* indicates that member A asked a question followed by member B responding, occurring 24 times across the dataset. The top 10 frequent sequences were visualized in Figure 3. Dyads in the Balanced pattern demonstrated symmetrical collaboration and mutual regulation. They frequently engaged in bidirectional actions such as *Question → Respond* exchanges and *Experiment → Experiment* transitions. Moreover, *A_Experiment → B_Share* and *B_Experiment → A_Organization* sequences reflected reciprocal knowledge co-construction and shared responsibility for task progression. In the Dominant pattern, one member (A) consistently initiated *Question–Experiment–Analyze*

 © ISLS

sequences, while the partner (B) mainly followed with *Experiment–Share*, indicating a leader–follower dynamic. The excerpt below illustrates a typical example of this pattern. In contrast, the Passive pattern exhibited sparse and fragmented interaction, dominated by *Experiment–Off-task* transitions and lacking other task-related behaviors. This suggests that these dyads were frequently disengaged from the inquiry process and demonstrated limited regulatory coordination.

Figure 3

The Visualization of Top Sequential Patterns of Each Pattern



Note. Ques=Question, Resp=Respond, Expt=Experiment, OffT=Off Task, Orgn = Organization

Group 6-A: Then how do we make orange? (Observe the experiment) (Question → Experiment)

Group 6-B: Maybe we can control the intensity? Brighten the red light. (Experiment → Share)

Group 6-A: (Operates the tablet) No, that's not working (Experiment → Analyze)

Group 6-B: Since it lets you click "+" or "–". (Experiment → Share)

Group 6-A: Oh! Nice! It worked!

Learning outcomes

A one-way ANOVA was conducted to compare students' scientific knowledge across the three patterns at both pre- and post-test. No significant differences were found at the pre-test, indicating comparable baseline knowledge levels. At the post-test, however, significant group differences emerged ($F = 5.03$, $p = .009$, $\eta 2 = 0.12$). Post-hoc comparisons revealed that students in the balanced (M = 71.47) and dominant patterns (M = 63.32) outperformed those in the passive pattern (M=59.17). No significant difference was observed between the balanced and dominant groups ($p > .05$). This suggests that active participation and task-focused collaboration, regardless of whether balanced or dominated, are associated with greater knowledge gains compared to disengaged collaboration.

Discussion and conclusion

This study identified three distinct role interaction patterns: Balanced, Dominant, and Passive, that emerged during AR-based collaborative inquiry. Dyads in the Balanced pattern exhibited high equality and mutuality, frequently engaging in bidirectional *Question–Respond–Experiment–Share* sequences and flexible role exchanges. The Dominant pattern showed moderate equality and mutuality, where one member consistently led task execution while the partner followed. Although these dyads displayed dominant roles and more unidirectional interactions, their outcomes were comparable to those of the Balanced pattern. This suggests that dominant or leader–follower structures do not necessarily impede learning. Prior work indicates that certain emergent roles can support productive collaboration by structuring interaction and maintaining task focus, even when participation is uneven (Saqr et al., 2024). For instance, in Group 6, both members contributed comparably through verbal exchanges, resulting in smooth collaboration. This pattern resembles the "expert–novice" structure, which, alongside the "collaborative" pattern, has been recognized as an effective form of dyadic interaction (Storch, 2002). Finally, Passive groups scored low on both dimensions, showing limited engagement and minimal task-related coordination, suggesting weak regulation and frequent disengagement. Future instructional research should focus on this pattern by designing scaffolds that promote meaningful interaction and sustained engagement among less active groups.

Overall, this study advances understanding of how role interaction unfolds in immersive learning environments and how role dynamics shape collaborative outcomes. The findings emphasize that balanced participation, mutual responsiveness, and dynamic role exchange are critical for fostering high-quality collaboration in AR-based inquiry. Although dominant patterns do not always impede learning, ensuring equitable participation remains essential. Future work should incorporate multimodal data (e.g., gestures, gaze, and speech) and explore adaptive scripting or AI-based scaffolds to support mutual regulation and prevent dominance or passivity in immersive collaborative learning environments.

References

Anna, G., & Spyros, V. (2025). Group dynamics in collaborative learning: Impact of emergent and scripted roles in tangible mobile augmented reality games. *Computers & Education: X Reality*, 7, 100102

Barron, B. (2003). When smart groups fail. *The Journal of the Learning Sciences, 12*(3), 307–359. https://doi.org/10.1207/S15327809JLS1203_1

Buchner, J., & Kerres, M. (2023). Media comparison studies dominate comparative research on augmented reality in education. *Computers & Education*, 195, 104711. https://doi.org/10.1016/j.compedu.2022.10471

Connor, C., Schoenborn, E. C., Hu, S., Porcino, T. M., Moore, C., Reilly, D., & Lages, W. S. (2024). *Examining pair dynamics in shared, co-located augmented reality narratives.* In *Proceedings of the ACM Symposium on Spatial User Interaction (SUI '24)* (pp. 1–11). ACM. https://doi.org/10.1145/3677386.3682091

Dowell, N., Lin, Y., Godfrey, A., & Brooks, C. (2020). Exploring the relationship between emergent sociocognitive roles, collaborative problem-solving skills, and outcomes: A group communication analysis. *Journal of Learning Analytics, 7*(1), 70–83. https://doi.org/10.18608/jla.2020.71.4

Dunleavy, M., Dede, C., & Mitchell, R. (2009). Affordances and limitations of immersive participatory augmented reality simulations for teaching and learning. *Journal of Science Education and Technology, 18*(1), 7–22. https://doi.org/10.1007/s10956-008-9119-1

Fischer, F., Kollar, I., Stegmann, K., & Wecker, C. (2013). Toward a Script Theory of Guidance in Computer-Supported Collaborative Learning. *Educational Psychologist, 48,* 56 - 66. https://doi.org/10.1080/00461520.2012.748005

Hou, H., & Wu, S. (2011). Analyzing the social knowledge construction behavioral patterns of an online synchronous collaborative discussion instructional activity using an instant messaging tool: A case study. *Computers & Education*, 57(2), 1459–1468.

Li, S., Jiao, X., Cai, S., & Shen, Y. (2025). Enhancing AR-based learning environments for STEM education: A design-based study on design features, kinematics learning and mathematics self-efficacy. *British Journal of Educational Technology*, 56(4), 1438-1462.

Jiao, X., Liu, Z., Zhou, H., & Cai, S. (2022). The effect of role assignment on students' collaborative inquiry-based learning in augmented reality environment. In *2022 International Conference on Advanced Learning Technologies (ICALT)* (pp. 349-351). IEEE

Mansour, N., Aras, C., Kleine Staarman, J., & Alotaibi, S. B. M. (2024). Embodied learning of science concepts through augmented reality technology. *Education and Information Technologies*, *29*, 1–31. https://doi.org/10.1007/s10639-024-12748-3

Radu, I., & Schneider, B. (2022). How augmented reality (AR) can help and hinder collaborative learning: A study of AR in electromagnetism education. *IEEE Transactions on Visualization and Computer Graphics, 29*(9), 3734–3745. https://doi.org/10.1109/TVCG.2022.3150489

Saqr, M., López-Pernas, S., & Murphy, K. (2024). How group structure, members' interactions and teacher facilitation explain the emergence of roles in collaborative learning. *Learning and Individual Differences.* https://doi.org/10.1016/j.lindif.2024.102463

Silva, M., Bermúdez, K., & Caro, K. (2023). Effect of an augmented reality app on academic achievement, motivation, and technology acceptance of university students of a chemistry course. *Computers & Education: X Reality, 2*, 100022. https://doi.org/10.1016/j.cexr.2023.100022

Srba, I., & Bielikova, M. (2014). Dynamic group formation as an approach to collaborative learning support. *IEEE transactions on learning technologies*, *8*(2), 173-186

Storch, N. (2002). Patterns of interaction in ESL pair work. *Language Learning*, 52(1), 119-158.

Upadhyay, B., Brady, C., Madathil, K. C., Bertrand, J., McNeese, N. J., & Gramopadhye, A. (2024). Collaborative augmented reality in higher education: A systematic review of effectiveness, outcomes, and challenges. *Applied ergonomics*, 121, 104360. doi:10.1016/j.apergo.2024.104360

Zhang, P., Li, J., & Cai, S. (2025). Timing matters: Effects of augmented reality game on students' learning achievement, satisfaction and acceptance. *British Journal of Educational Technology, 56*(3), 1273–1293. https://doi.org/10.1111/bjet.13524

International Society of the Learning Sciences

Inquiry Script Modeling: A Formalism for Scripting and Orchestration of Collective Inquiry

Erfane Ghasempour Khoshroodi, Jim Slotta
Erfane68@gmail.com, jslotta@gmail.com
University of Toronto

Abstract: This paper introduces Inquiry Script Modeling (ISM), a formalism designed to represent and operationalize inquiry-oriented learning workflows in classrooms that adopt a learning community approach. Drawing on prior CSCL research, including process modeling and orchestration frameworks, ISM supports interpretability, temporal sequencing, and adaptability. This paper presents ISM using a use case focused on AI and disinformation, to show how such formalisms can help teachers to design, communicate, and dynamically manage complex collaborative learning scripts.

Keywords: Scripting, Orchestration, Collective Inquiry, Formalism

Introduction

Standardized formalisms such as chemical notation or calculus, accelerate scientific progress by enabling clear theory-building and collaborative discourse (Abe, 2012). The learning sciences, however, lack such universal representations for pedagogical designs. While CSCL "scripts" define learner interactions and roles (Kollar et al., 2006), and frameworks like IMS-LD attempted to standardize these processes across contexts (Koper & Olivier, 2004), a gap remains in modeling guidance and orchestration across diverse activity types (Fischer & Slotta, 2013).

The quest to formalize learning processes has consistently struggled to balance computational rigor with conceptual clarity. Early standards like IMS-LD attempted platform-agnostic modeling but were often critiqued for technical complexity and an inability to capture the dynamic, social nature of collaboration (Koper & Tattersall, 2005; Hernández-Leo et al., 2005). This highlights a persistent challenge: opaque formalisms offer limited utility if they cannot be interpreted by the stakeholders they are meant to serve. While CSCL frameworks - from micro-level interaction moves to macro-level orchestration - have defined essential building blocks like roles and resources (Harrer et al., 2007; Dillenbourg, 2013), these representations are never truly "agnostic." Each carries epistemic commitments that prioritize specific pedagogical values, such as rigid task sequencing over adaptive roles. Even breakthroughs like Orchestration Graphs often remain too coarse for complex activity designs. Shifting toward a Learning Communities framework, our work develops granular, machine-readable scripts that explicitly reflect the theoretical commitments of collective inquiry. This structured approach aims to enhance the shareability and execution of designs while safeguarding human agency and addressing the orchestrational demands of technology-enhanced environments.

The present work begins with the premise that educators, including researchers and practitioners, will benefit from representational tools that are both formal and usable: formal enough to be machine-readable and support computational orchestration, yet usable enough to be meaningful and adaptable for educators in practice. In response to this fundamental tension, we have developed Inquiry Script Modeling (ISM) as a formalism that enables educators (including researchers, who are an important user group for ISM) to specify inquiry-oriented learning designs, communicate them effectively to other practitioners, researchers or system designers, and adapt them dialogically during the design process. Our guiding research question is: What elements should be included within a scripting formalism to enable educational researchers and practitioners to communicate about, and co-develop curricular designs, relating to inquiry-oriented pedagogy?

Methodological approach

Using a Design-Based Research (DBR) methodology, this study details the iterative development of the *Inquiry Script Modeling (ISM)* as both a theoretical outcome and a functional tool for computational orchestration (Barab & Squire, 2004). A multidisciplinary team of learning scientists, AI researchers, and software engineers conducted two primary design cycles. The first cycle involved a comparative analysis of CSCL scripting, Petri Nets, and Directed Acyclic Graphs (DAGs) to identify core modeling primitives (e.g., control flow, transitions, and temporal

 © ISLS

sequencing). In the second cycle, these constructs were operationalized through an authentic inquiry use case regarding misinformation, leading to the refined node taxonomies and AI-supported orchestration principles presented in Figure 1 and Table 2. This DBR phase establishes ISM's conceptual foundation, derived from systematic analysis and iterative refinement.

Results and analysis

An analysis of core functions across formalisms

To inform the development of *ISM*, we analyzed five formalisms across computer science, business, and the learning sciences: *Process Modeling, Petri Nets, Directed Acyclic Graphs (DAGs), Orchestration Graphs, and CSCL scripting models.* As shown in Table 1, these frameworks offer complementary strengths for modeling complex workflows but vary significantly in their pedagogical and computational utility.

Petri Nets and DAGs provide mathematical rigor for formalizing control flow, concurrency, and dependencies, making them ideal for scalable, machine-executable workflows. However, they lack intrinsic support for *pedagogical semantics*, such as social roles, adaptive feedback, or learner agency. Conversely, *Orchestration Graphs and CSCL scripts* are highly expressive of pedagogical intent and social-plane interactions (individual, group, and class levels) but often lack the structural constraints required for direct computational execution (Dillenbourg, 2013; Harrer et al., 2007). Our analysis suggests that while the CSCL community has proposed tailored languages (e.g., Miao et al., 2005), limited uptake persists due to the difficulty of balancing *human readability with machine interpretability.* ISM is designed to bridge this gap, synthesizing the structural clarity of Petri Nets and DAGs with the social coordination and instructional richness of Orchestration Graphs.

Table 1

A Comparative Analysis of Key Features in Orchestration Graphs, Petri Nets, and DAGs

Function	Orchestration Graphs	Petri Net	DAG
Nodes	Learning activities, teacher/student actions	Places (states), Transitions (events)	Tasks, events, or data points
Edges	Dependencies; represent pedagogical, computational, or control-flow relations; edges are directed, weighted, elastic (decay over time)	Directed edges showing token flow between places and transitions	Directed edges indicating dependency/precedence
Transitions	Probabilistic state transitions	Explicit transitions representing events/actions	implicit; tasks complete when all dependencies are satisfied.
Data Flow	Managed via operators; move or transform data between activities (e.g., storing, sharing, or distributing learner.	Tokens represent movement of resources, data, or work items	Explicit; edges can carry or annotate data dependencies; metadata can augment flow analysis

The ISM formalism: visualization

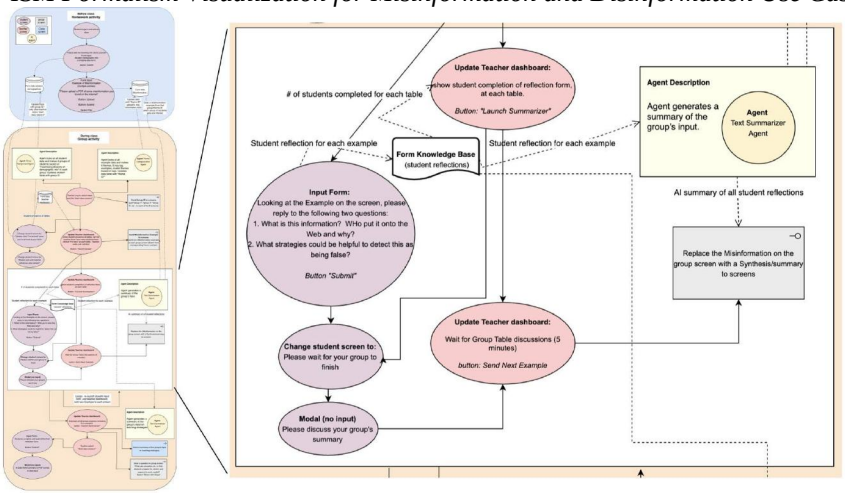
The ISM formalism has been visualized as a multi-layered, color-coded flowchart that differentiates actors, data, and knowledge artifacts while explicitly representing AI and instructor orchestration. The design emphasizes interpretability, real-time adaptability, and alignment with established CSCL scripting principles. The table 2 summarizes the core elements of the ISM formalism, detailing how nodes, edges, transitions, and flows are represented. It highlights how students, teachers, and AI agents interact, how data and control flows are managed, and how temporal sequencing is visualized within inquiry-oriented learning activities.

 © ISLS

Table 2
ISM Formalism Functions and Elements

Function	ISM Elements
Nodes	Student screens (purple ovals), teacher screens (pink ovals), AI agents (yellow circles), input forms (ellipses), data tables (cylinders), knowledge base (bold rectangles)
Edges	Solid arrows for control flow (task sequencing, teacher-mediated transitions); dashed arrows for data flow (movement of inputs, outputs, and processed knowledge)
Transitions	Labeled to indicate conditions, actions, or branching logic; iteration loops explicitly marked for repeated processes
Data Flow	Dashed arrows move raw inputs (demographics, uploaded examples) to AI or teacher processing and transform into aggregated knowledge in the knowledge base
Control Flow	Solid arrows guide task progression and sequencing; teacher retains agency over sequencing, while AI automates task distribution and group formation
Temporal Representation	Vertical layout represents chronological order (Before Class → During Class), showing stepwise progression and interaction timing

Figure 1
ISM Formalism Visualization for Misinformation and Disinformation Use Case



Use case: Disinformation and misinformation in learning spaces

To illustrate Inquiry Script Modeling (ISM), we modeled a collaborative inquiry into misinformation where students analyze real-world cases to develop response strategies. The script progresses through four AI-coordinated stages: (1) *Pre-class*, where students upload examples and demographics for AI-driven categorization and group formation; (2*) In-class Group Analysis*, a 3–4 cycle loop where the system pushes case studies to group screens and individual devices, followed by AI-generated syntheses for discussion; (3) *Class Sharing*, where the AI clusters group strategies into thematic categories for a live, class-wide debrief; and (4) *Post-class*, where students receive an automated PDF summary of the community's collective strategies upon submitting reflection forms. This scenario demonstrates ISM's capacity to formalize complex data flows and AI-mediated coordination within a learning community.

 © ISLS

ISM: design considerations

The Inquiry Script Modeling (ISM) formalism is built upon five foundational design pillars that balance computational rigor with pedagogical usability. *Balancing Simplicity and Complexity* navigates the trade-off where detailed models reveal essential workflows but risk overwhelming users while overly simple diagrams may miss key teaching or timing relationships. ISM addresses this through selective abstraction and visual diversity, leveraging Colored Petri Nets and Business Process Model notation with varied shapes and colors to distinguish actors and activities. Purple ovals represent students, pink ovals represent teachers, and yellow circles represent AI agents. Solid lines denote control flow while dashed lines denote data flow. Vertical ordering establishes temporal sequencing and labels clarify logic for actions and transitions, ensuring the visualization captures complex processes while remaining interpretable and actionable (Möller et al., 2025). *Temporal Sequencing and Flow Distinction* is critical for effective orchestration, requiring a clear separation of control flow (the sequence and timing of actions) from data flow (the movement and transformation of learning artifacts into shared knowledge). Because educational workflows involve complex dependencies between human and computational actors, ISM provides teachers a "mastery" of timing by seeing real-time statuses of student devices and ongoing activities (Dillenbourg & Jermann, 2010). This ensures interventions occur at the right moment to support engagement. ISM utilizes AI agents to automate high-load tasks such as analyzing submitted examples to generate thematic categories, tagging individual data points, and updating central tables with a "theme ID." This automation reduces teacher cognitive load and ensures timely distribution of content without creating bottlenecks (Bergmann & Gil, 2014). *Flexibility and Adaptability* requires balancing dynamic orchestration with pedagogical coherence (Dillenbourg & Jermann, 2010). Our design minimizes extrinsic constraints originating from technical limitations while preserving intrinsic constraints or core pedagogical rules to ensure instructional intent is maintained (Dillenbourg & Tchounikine, 2007). In the ISM framework, AI agents handle repetitive and computationally intensive tasks such as data clustering and content delivery while teachers retain "driver" status, controlling sequencing and timing to avoid the pitfalls of overscripting (Harrer & Lingnau, 2016). This hybrid approach ensures adaptability, allowing teachers to take ownership of the script and adjust it based on evolving classroom contexts (Tchounikine, 2011). For instance, AI agents notify teachers of student progress but the human instructor decides when to advance the activity. *Making Pedagogical Topologies Visible* addresses the sociomaterial connections among users, devices, displays, and content (Castañeda et al., 2024). ISM clarifies this ontology by differentiating between personal student devices and shared group displays while labeling the specific content such as a "form" or "page" on each. This dual-layer visibility allows teachers to detect participation bottlenecks or synchronization mismatches such as a group display showing outdated content before a result appears. This enhances teacher agency by enabling targeted adjustments like redistributing content or shifting activities between device types (Liu et al., 2023). *Integrating Physical, Social, and Curricular Spaces* leverages the HCI concept of the ambient interface to extend the user interface to the whole environment. By orchestrating across individual learner devices, group collaboration at tables, and whole-class activities mediated by large shared displays, ISM captures how the curricular space (learning objectives and inquiry tasks) intersects with physical devices and social interactions. This layered representation provides a coherent framework for embodied, situated interaction within a collective inquiry context.

Conclusion

This paper has examined the challenges and opportunities of formalizing inquiry-oriented learning designs, with a focus on learning communities frameworks. By reviewing prior formalisms from learning sciences, computer science, and business process management, we highlighted the trade-offs between formal rigor, interpretability, and pedagogical usability. Our proposed ISM formalism demonstrates how complex learning workflows can be made visible, interpretable, and actionable for both educators and computational systems. A key contribution of this work is its theoretical saliency: ISM makes learning communities elements, knowledge building, inquiry scaffolding, and shared knowledge artifacts, visible, allowing designers and teachers to reason about and reflect on how these elements are enacted in practice (Peters, 2011; Acosta & Slotta, 2018). It is important to note that at this stage, ISM is still in the design and development phase. We do not yet offer a fully machine-readable implementation; the formalism remains under active research, focusing on interpretability, usability, and the visualization of learning communities elements.

References

Abe, T. O. (2012). Genesis of Selected Mathematical Notations. Journal of Research in Education and Society (JRES) , 3(2), 24–28.

Acosta, A., & Slotta, J. D. (2018). Scripted Idea Improvement in a Learning Community Curriculum for Grade 12 Biology. Conference Paper, Knowledge Building Summer Institute.

Barab, S., & Squire, K. (2004). Design-Based Research: Putting a Stake in the Ground. Journal of the Learning Sciences, 13(1), 1–14. https://doi.org/10.1207/s15327809jls1301_1

Bergmann, R., & Gil, Y. (2014). Similarity assessment and efficient retrieval of semantic workflows. Information Systems, 40, 115-127.

Dillenbourg, P., & Jermann, P. (2010). Technology for classroom orchestration. In New science of learning: Cognition, computers and collaboration in education (pp. 525-552). New York, NY: Springer New York.

Dillenbourg, P., & Tchounikine, P. (2007). Flexibility in macro CSCL scripts. Journal of Computer Assisted Learning, 23(1), 1–13.

Dillenbourg, P. (2013). Design for classroom orchestration. Computers & education, 69, 485-492.

Fischer, F., Slotta, J. D., Dillenbourg, P., Kollar, I., Stegmann, K., Wecker, C., & Tchounikine, P. (2013). Scripting and orchestration: recent theoretical advances.

Harrer, A., Kobbe, L., & Malzahn, N. (2007). Conceptual and computational issues in the formalization of collaboration scripts. In C. A. Chinn, G. Erkens, & S. Puntambekar (Eds.), Proceedings of the 8th International Conference on Computer Supported Collaborative Learning (pp. 280–282). International Society of the Learning Sciences.

Harrer, A., & Lingnau, A. (2016, July). Bringing Balance to the Forces of Learning Design: Integrating Self-Regulated Planning and Scripting via Model Transformations. In 2016 IEEE 16th International Conference on Advanced Learning Technologies (ICALT) (pp. 158-162). IEEE.

Hernández-Leo, Davinia, Juan I. Asensio-Pérez, Yannis Dimitriadis, Miguel L. Bote-Lorenzo, Iván M. Jorrín-Abellán, and Eloy D. Villasclaras-Fernández. "Reusing IMS-LD formalized best practices in collaborative learning structuring." (2005).

Kollar, I., Fischer, F., & Hesse, F. W. (2006). Collaboration scripts–a conceptual analysis. Educational Psychology Review, 18(2), 159-185.

Koper, R., & Olivier, B. (2004). Representing the learning design of units of learning. Journal of Educational Technology & Society, 7(3), 97–111.

Liu, Q., Gladman, T., Muir, J., Wang, C., & Grainger, R. (2023). Analytics-informed design: Exploring visualization of learning management systems recorded data for learning design. SAGE Open, 13(3), 21582440231193590.

Miao, Y., Hoeksema, K., Hoppe, H. U., & Harrer, A. (2005). CSCL scripts: Modelling features and potential use. In Proceedings of the 2005 Conference on Computer Supported Collaborative Learning (pp. 423–432). International Society of the Learning Sciences.

Möller, M., Winter, M., & Reichert, M. (2025). Cognitive factors in process model comprehension—A systematic literature review. Brain Sciences, 15(5), 505. https://doi.org/10.3390/brainsci15050505

Peters, V. L. (2011). Knowledge Community and Inquiry in Secondary School Science (Unpublished doctoral dissertation). University of Toronto.

Tchounikine, P. (2011). Computer science and educational software design: A resource for multidisciplinary work in technology enhanced learning (pp. I-XII). Berlin: Springer.

Tissenbaum, M., & Slotta, J. D. (2013). Scripting and orchestration of collaborative inquiry: An increasing complexity of designs. International Society of the Learning Sciences.

Weinberger, A., Stegmann, K., Fischer, F., & Mandl, H. (2007). Scripting argumentative knowledge construction in computer-supported learning environments. In Scripting computer-supported collaborative learning: Cognitive, computational and educational perspectives (pp. 191-211). Boston, MA: Springer US.

International Society of the Learning Sciences

The Effects of GenAI-Assisted Programming Learning on Fifth Graders' Computational Thinking, Programming Self-Efficacy and Outcomes, and Programming-oriented Prompting: A Comparison of Two Human–AI Collaboration Configurations

Ying-Tien Wu, Yi-Hong Zeng
ytwu@cl.ncu.edu.tw, beliefloatsky@gmail.com
National Central University, Taiwan

Ching-Hui Kuan, Lilin Elementary School, Taiwan, arielkch1127@gmail.com

Abstract: This study investigates the effects of GenAI-assisted programming learning on fifth graders' computational thinking, programming self-efficacy and outcomes, and programming-oriented prompting. A total of 99 fifth graders from four classes participated and were randomly assigned to two GenAI-assisted programming instruction groups with different human–AI collaboration configurations: single student-AI collaboration (SSAC), in which each student worked individually with a computer with ChatGPT, and dyad student-AI collaboration (DSAC), in which two students worked in pairs, collaboratively using a shared computer with ChatGPT. The instruction was implemented over 6 weeks (1 class per week). This study revealed no significant difference in computational thinking between the two groups of students after GenAI-assisted programming instruction ($p>0.05$). In contrast, the students in the DSAC group showed significantly higher programming self-efficacy and better programming outcomes ($p<0.05$). Moreover, significant differences in the distribution of programming-oriented prompting between these two groups of students were also found ($p<0.05$).

Introduction

In recent years, programming education for younger students has been highlighted (Lai et al., 2021). Conventionally, programming consists of three major iterative phases, including design, implementation, and debugging. In the design phase, learners specify the needs, requirements and constraints of an application, online system or platform, breaking it down into smaller, more manageable subtasks. During the implementation phase, the requirements are translated into executable code to build a working prototype. In the debugging phase, the prototype is evaluated and, through processes such as algorithmic improvements and the identification of syntactic errors, ultimately optimized (Abouelenein et al., 2025). In this context, computational thinking, encompassing four key components, including decomposition, pattern recognition, abstraction and algorithm design, is widely recognized as crucial for programming (Ogegbo & Ramnarain, 2022).

With visual and block-based coding languages and tools, such as Scratch, students have increasing opportunities to engage in programming at an earlier age in elementary schools (e.g., Xu et al., 2023). These virtual programming environments reduce the syntactic complexity of programming, allowing beginners to focus more on the design stage (Wong, 2023). Previous research has also confirmed that visual programming environments can enhance students' positive attitudes and their willingness to engage in solving complex problems (Montuori et al., 2024). Although visual programming environments are indeed a promising entry point, many elementary school students still have difficulty learning abstract programming concepts and computational thinking (Grover, 2021). In particular, these visual programming environments are still insufficient for younger students to develop computational thinking meaningfully.

In recent years, with the emergence of generative AI (GenAI), its potential in AI-assisted programming has been highlighted (Liu et al., 2026). Because of its accessibility, students can program in natural language, making programming more intuitive. In particular, as advocated by researchers, GenAI, such as ChatGPT and other large language models (LLMs), has much potential in developing elementary students' computational thinking (CT) (Yilmaz & Karaoglan Yilmaz, 2023). However, empirical research examining this perspective remains limited. Moreover, as advocated in previous studies, in early programming education, student self-efficacy towards programming may play a more critical role than learning outcomes or performance because confidence can significantly affect continuous programming learning (e.g., Xia et al., 2022). That is, GenAI-assisted programming experiences are likely to enhance young students' self-efficacy toward programming from the outset. Also, the effects of AI-assisted programming on young students' programming self-efficacy (PE) are underexplored.

 © ISLS

To address the aforementioned important issues, the current study examined how GenAI-assisted programming learning affects younger students' programming learning. In particular, two human–AI collaboration configurations, single student-AI collaboration (SSAC) and dyad student-AI collaboration (DSAC), were developed and integrated into GenAI-assisted programming learning activities in this study. Then, the effects of GenAI programming learning activities, under the two human–AI collaboration configurations, on elementary school students' computational thinking (CT), programming self-efficacy (PE), and programming outcomes were examined separately and compared across groups.

The research questions of this study are as follows: (1) Do GenAI-assisted collaborative programming activities significantly affect elementary students' CT and PE? (2) After experiencing GenAI-assisted programming activities, do the elementary students with different human–AI collaboration configurations have any significant difference in their CT, PE and programming outcomes? (3) What are the elementary students' programming-oriented prompting revealed in GenAI-assisted programming? Do elementary students with different human–AI collaboration configurations show significant differences in their programming-oriented prompting?

Method

This study was conducted with a quasi-experimental design, with nonequivalent groups and a pretest–posttest design. A total of 99 fifth graders from four classes at a public elementary school in Taiwan participated in this study. These classes were randomly assigned to two instructional intervention (i.e., GenAI-assisted programming activities) groups with different human–AI collaboration configurations: single student-AI collaboration (SSAC), comprising two classes, and dyad student-AI collaboration (DSAC), comprising two classes.

The primary goal of GenAI-assisted programming activities in this study was to develop a web-based game through collaboration with GenAI. The activities were implemented over 6 weeks, with 1 class (40 minutes) per week (240 minutes in total). In the first half of the course, each student used a computer to build fundamental programming skills in collaboration with GAI. In the latter half of the course, students were assigned to two different human–AI collaboration configuration groups, SSAC and DSAC. In the SSAC group, each student worked individually with a computer with a GenAI application (i.e., ChatGPT) to develop a web-based game, whereas in the DSAC group, two students worked in pairs, collaboratively using a shared computer with a GenAI application (i.e., ChatGPT) to develop a web-based game.

To evaluate students' CT skills, PE, and the quality of their programming outcomes (i.e., the web-based game), three instruments were used in this study. The Computational Thinking Scale, adapted from Korkmaz et al. (2017), was used to assess the students' CT. The scale comprised five subfactors: creativity, algorithmic thinking, cooperativity, critical thinking, and problem-solving, with 7, 6, 4, 5, and 6 items, respectively, for a total of 28 items. The overall reliability of the scale was 0.84, revealing high reliability for assessing students' computational thinking. The Programming Self-Efficacy Scale, adapted from Kukul et al. (2017), was used in this study to evaluate students' programming self-efficacy. It is a unidimensional scale comprising 31 items. The overall reliability of the scale was calculated to be 0.95. Moreover, the Game Design Rubric was used to evaluate the students' programming outcome in this study. It was adapted from a previous study (M.I et al., 2021), which comprises four dimensions: originality, fluency, flexibility and elaboration ($\alpha = .88$). In this study, the students' programming outcomes (i.e. their web-based game) were scored from 1 to 3 points for each dimension.

In addition, students' prompts used for automatic code generation with ChatGPT were analyzed using a three-layer coding scheme. Typically, programming involves three major phases: design, implementation and debugging. Accordingly, this study identified three types of programming-oriented prompting with GenAI: design-oriented prompting, focusing on creating or modifying game features; coding-oriented prompting, referring to direct code-generation requests with minimal elaboration; debugging-oriented prompting, concerning the identification and correction of program errors. Moreover, both the design-oriented and debugging-oriented promptings were further categorized into two subcategories: UI (interface)-related prompting and UX (game element)-related prompting.

Major findings

Effects of AI-assisted programming activities

In this study, the effects of GenAI-assisted programming activities using the single student–AI collaboration (SSAC) and dyadic student–AI collaboration (DSAC) configurations on fifth-graders' CT and PE were examined. As shown in Table 1, the single student-AI collaboration (SSAC) group showed no significant differences in students' CT and PE between the pretest and the post-test. However, different results were found in the dyad student-AI collaboration (DSAC) configuration group. In particular, in the DSAC group, a significant

improvement in students' PE was revealed between the pre-test and the post-test ($p < .05$), whereas the CT was not significant. It indicated that only GenAI-assisted programming with the dyad student-AI collaboration (DSAC) configuration could enhance students' confidence in programming. One possible explanation is that collaboration with a peer allows students to seek support and exchange ideas, which in turn fosters a greater sense of competence during the activities. However, no significant improvement in CT was observed in either group following the GenAI-assisted programming activities. It suggests that further instruction in computational thinking, such as explicit CT skills instruction or guided, deliberate reflection, is needed and should be integrated into GenAI-assisted programming learning.

Table 1

Pre- and Posttest Scores of CT and PE Across Collaboration Configurations

Variable	Group	Pretest M (SD)	Posttest M (SD)	*t*
CT	SSAC	94.68 (18.10)	93.94 (16.36)	-0.33
	DSAC	97.82 (19.63)	94.56 (19.96)	-1.31
PE	SSAC	99.72 (31.31)	103.34 (21.51)	0.63
	DSAC	105.48 (34.64)	119.88 (22.49)	2.59*

*$p<.05$

Comparisons of students' computational thinking and programming self-efficacy between the two different human–AI collaboration configuration groups

The ANCOVA results, as shown in Table 2, revealed no significant difference in CT between the DSAC and SSAC groups, indicating that the two groups performed similarly in CT after adjusting for pretest scores. In contrast, for programming self-efficacy (PE), the DSAC group outperformed the SSAC group ($p < .001$), suggesting that DSAC can significantly enhance elementary students' confidence in programming compared with SSAC. It suggests that pairing students with GenAI in dyads may be a promising strategy to build programming confidence.

Table 2

ANCOVA Results for CT and Programming Self-Efficacy (PE) Across Collaboration Configurations

Variable	Group Type	Adjusted Mean	F(1, 96)	p	Partial η^2
CT	SSAC	94.79	0.16	.687	.002
	DSAC	93.59			
PE	SSAC	103.03	14.16	< .001**	.129
	DSAC	*119.69*			

**$p<.01$

Comparisons of students' programming outcome between the two different human–AI collaboration configuration groups

Table 3 revealed that the students in the DSAC group significantly outperformed their counterparts in all four dimensions of programming outcome (i.e., game design) ($p < .05$), indicating that, compared with single student-AI collaboration, dyad student-AI collaboration could significantly improve elementary students' programming outcome in terms of the originality, the fluency, the flexibility and elaboration of their game design.

Table 3

Descriptive Statistics of Game Creativity Scores by Collaboration Configuration

Dimension	SSAC M (SD)	DSAC M (SD)	F(1,97)	p	η^2
Originality	1.20 (0.46)	1.98 (0.75)	34.64	< .001***	.300
Fluency	1.34 (0.44)	2.02 (0.52)	39.28	< .001***	.327
Flexibility	1.41 (0.54)	1.91 (0.65)	13.97	< .001***	.147
Elaboration	1.44 (0.50)	1.73 (0.70)	4.91	.029*	.057

*$p<.05$; ***$p<.001$

 © ISLS

Patterns of students' programming-oriented prompting between the two different human–AI collaboration configuration groups

A chi-square test was performed to examine whether the two groups of students differed in their proportions of the two prompting: Design-oriented prompting–UX + Debugging-oriented prompting–UX and Design-oriented prompting–UI + Debugging-oriented prompting–UI. As shown in Table 4, a significant difference was found (p<0.01). The students in the DSSC group were more inclined to engage in UX-related prompting (74.1%), focusing on game rules and logic, whereas those in the one-student-one-machine group concentrated more on interface design and visual adjustments (59.0%).

Table 4

Distribution of Interaction Types across Collaboration Configurations

	SSAC *(n = 346)*	DSAC *(n = 147)*
Design-based prompting–UI + Debugging–UI	142 (41.0%)	38 (25.9%)
Design-based coding–UX + Debugging–UX	204 (59.0%)	109 (74.1%)
χ^2	10.27**	

**p<.01

As revealed in the aforementioned discussion, the students in DSAC showed higher programming self-efficacy and better programming outcomes. Moreover, they also had relatively higher UX-related prompting. It seems that their higher UX-related prompting may be related to their more sophisticated game design outcomes and higher programming self-efficacy. However, further investigation is needed to provide insights into this issue.

Conclusion

In conclusion, two human-AI collaboration configurations, single student-AI collaboration (SSAC) and dyad student-AI collaboration (DSAC), were developed and integrated into GenAI-assisted programming activities in this study. It was revealed that, compared to SSAC, DSAC could significantly improve elementary students' confidence in programming, which may play a critical role in their continuing programming learning. Similar findings were also found in students' AI-assisted programming outcomes. More importantly, computational thinking did not improve significantly in either group, suggesting that further efforts are needed to develop effective scaffolding to support the development of students' computational thinking in GenAI-assisted programming environments.

References

Abouelenein, Y. A. M., Ghazala, A. F. A., Mahdy, E. M. M., & Khalaf, M. H. R. (2025). The R5E pattern: can artificial intelligence enhance programming skills development? *Education and Information Technologies*, *30*(15), 22177-22205. https://doi.org/10.1007/s10639-025-13616-3

Grover, S. (2021). Teaching and Assessing for Transfer from Block-to-Text Programming in Middle School Computer Science. In. https://doi.org/10.1007/978-3-030-65632-4_11

Korkmaz, Ö., Çakir, R., & Özden, M. Y. (2017). A validity and reliability study of the computational thinking scales (CTS). *Computers in Human Behavior*, *72*, 558-569. https://doi.org/https://doi.org/10.1016/j.chb.2017.01.005

Kukul, V., Gökçearslan, Ş., & Günbatar, M. S. (2017). Computer programming self-efficacy scale (CPSES) for secondary school students: Development, validation and reliability. *Educational Technology Theory and Practice*, *7*, 158-179. https://doi.org/10.17943/ETKU.72918

Lai, C.-F., Zhong, H.-X., & Chiu, P.-S. (2021). Investigating the impact of a flipped programming course using the DT-CDIO approach. *Computers & Education*, *173*, 104287. https://doi.org/https://doi.org/10.1016/j.compedu.2021.104287

Liu, J., Zhang, Y., Li, W., Wang, Q., Niu, P., & Zhang, X. (2026). Adaptive vs. planned metacognitive scaffolding for computational thinking: Evidence from generative AI-supported programming in elementary education. *Computers & Education*, *241*, 105473. https://doi.org/https://doi.org/10.1016/j.compedu.2025.105473

M.I, L. F., Yatim, M., Mohamed Zain, N. Z., Azizan, U. H., & Fabil, N. (2021). Development of Rubric to Measure Children's Creativity in Game Design. *The International journal of Multimedia & Its Applications*, *13*, 27-34. https://doi.org/10.5121/ijma.2021.13203

Montuori, C., Gambarota, F., Altoé, G., & Arfé, B. (2024). The cognitive effects of computational thinking: A systematic review and meta-analytic study. *Computers & Education*, *210*, 104961. https://doi.org/https://doi.org/10.1016/j.compedu.2023.104961

Ogegbo, A. A., & Ramnarain, U. (2022). A systematic review of computational thinking in science classrooms. *Studies in Science Education*, *58*(2), 203-230. https://doi.org/10.1080/03057267.2021.1963580

Wong, G. (2023). Amplifying children's computational problem-solving skills: A hybrid-based design for programming education. *Education and Information Technologies*, *29*, 1-33. https://doi.org/10.1007/s10639-023-11880-9

Xia, Q., Chiu, T. K. F., Lee, M., Sanusi, I. T., Dai, Y., & Chai, C. S. (2022). A self-determination theory (SDT) design approach for inclusive and diverse artificial intelligence (AI) education. *Computers & Education*, *189*, 104582. https://doi.org/https://doi.org/10.1016/j.compedu.2022.104582

Xu, E., Wang, W., & Wang, Q. (2023). A meta-analysis of the effectiveness of programming teaching in promoting K-12 students' computational thinking. *Education and Information Technologies*, *28*(6), 6619-6644. https://doi.org/10.1007/s10639-022-11445-2

Yilmaz, R., & Karaoglan Yilmaz, F. G. (2023). The effect of generative artificial intelligence (AI)-based tool use on students' computational thinking skills, programming self-efficacy and motivation. *Computers and Education: Artificial Intelligence*, *4*, 100147. https://doi.org/https://doi.org/10.1016/j.caeai.2023.100147

Acknowledgments

This research was supported by the National Science and Technology Council (NSTC), Taiwan (Grant numbers: 112-2410-H-008-036-MY3 and 114-2423-H-008-002-). The authors gratefully acknowledge the financial support provided.

International Society of the Learning Sciences

STEM Teachers' Perspectives on Generative AI: A Social Media Content Analysis of Reddit Discussions

Leqi Li, Pennsylvania State University, lzl5599@psu.edu
Xin Guan, Columbia University, xg2413@tc.columbia.edu

Abstract: The rapid emergence of generative AI (GenAI) tools has sparked both optimism and skepticism among STEM educators. While prior research has examined teacher perceptions through surveys, interviews, and adoption frameworks, less is known about educators' large-scale, spontaneous, and grassroots discourse. This study analyzes STEM teachers' discussions about GenAI on Reddit to understand how teachers collectively construct perspectives on its classroom use. Using BERTopic for topic modeling and an adjusted Technology Acceptance Model (TAM) framework, we examined posts and comments from educator-focused subreddits. Results revealed nine dominant themes, highlighting tensions around academic integrity and ethical student use, alongside pragmatic uses such as lesson planning and grading support. Attitudes were the most frequently coded dimension, whereas perceived ease of use was less salient. These findings underscore teachers' cautious stance toward GenAI, emphasizing the need for professional development, AI literacy, and policies that support critical and equitable integration.

Introduction

The application of Generative AI (GenAI) in STEM education has been growing because of its potential to transform teaching and learning. As GenAI tools begin to appear in classrooms, teachers' perspectives play a key role in shaping how these technologies are adopted in practice. It is crucial to examine how teachers perceive, accept, and engage with these technologies to ensure that GenAI supports, rather than replaces, teachers' agency and autonomy (Collie & Martin, 2024). Existing studies have examined STEM teachers' perceptions from diverse angles, including epistemic cognition (Garofalo & Farenga, 2025), technology adoption frameworks (Zhai & Krajcik, 2024), targeted interventions (Kim, 2024), and sociotechnical perspectives (Cheah & Kim, 2026).

Among multiple perspectives, the Technology Acceptance Model (TAM) framework has been widely used to explore teachers' perceptions and attitudes toward generative AI. As one of the technology adoption models, TAM explains an individual's technology adoption based on the core dimensions of perceived usefulness (PU), perceived ease of use (PEU), and attitude towards using (ATT) that influence the behavioural intention (BI) on the technology (Davis et al., 1989). In educational research, it is an influential model adapted and validated across various contexts and technologies to explore the acceptance of educational tools from teachers' perspectives (Salahshour Rad et al., 2018; Zhai, 2025). With the increasing integration of GenAI into classrooms, a growing body of research has employed and extended the TAM frameworks to explore teachers' perspectives in response to the contextual complexities of these emergent technologies (Hsiao & Tang, 2025; Kong et al., 2024; Nevárez Montes & Elizondo-Garcia, 2025). For instance, Kong et al. (2024) proposed an extended TAM to investigate K-12 teachers' acceptance of GenAI. It is based on the integrated TAM, which incorporates additional dimensions of self-efficacy (SE) and subjective norm (SN) (Davis et al., 1989; Venkatesh & Davis, 2000), to better understand how teachers' adoption of GenAI tools is influenced both internally and externally. The current study adopts this extended version of the TAM as an interpretive framework to investigate STEM teachers' perceptions, opinions, and discourse towards GenAI.

While prior work has employed surveys and semi-structured interviews to collect teachers' perceptions of GenAI tools, social media platforms such as X and Reddit provide additional complementary sources that capture educators' opinions about GenAI in education (Smith-Mutegi et al., 2025). These open, data-rich forums provide informal and anonymous spaces that encourage greater self-disclosure, especially for intimate or negative content (Ma et al., 2016). Due to the lack of data-based research about teachers' diverse perspectives on AI in STEM education on social media, the current study aims to analyze STEM teachers' discussions about GenAI in Reddit subgroups to examine how educators collectively construct their perspectives on GenAI. Specifically, the study aims to explore the *Research Question*: What themes characterize STEM teachers' discussions of generative AI in online communities? To address this question, we apply BERTopic for topic modeling and analyze the discourse with the adjusted TAM framework to identify key themes and patterns.

Method

 © ISLS

Reddit, a social media platform that researchers have utilized to examine specific themes within educational communities (e.g., Wu et al., 2024), is selected as the data collection site. Data collection was conducted using the Python Reddit API Wrapper (PRAW), capturing the data from October 2023 to October 2025. We authenticated through a research account to query the subreddits about posts and comments, which are: r/Teachers (2.2M visitors), r/education (218k), r/edtech (31k), and r/Professors (167k). For each subreddit, we retrieved the up to 500 most recent and high-upvote posts, capturing the generative AI-related threads among the STEM teacher community. With the search strategy, we got 213 posts with 2057 corresponding comments.

We combined the posts and the comments into a dataset as our corpus. After pre-processing, we applied the BERTopic algorithm, which integrates vectorization, dimensionality reduction, and clustering to generate topics. BERTopic is a topic modeling method that employs transformers and c-TF-IDF to create interpretable topics (Grootendorst, 2022) and is suitable for short and informal texts on social media (Egger & Yu, 2022). We also finetuned the model by applying UMAP to reduce dimensionality while preserving local semantic neighbors, followed by HDBSCAN to identify clusters suitable for short and sparse online discourse. The final model yielded 29 topics, each defined by representative keywords and example posts/comments. After manual checking and synthesizing, we report the most dominant 9 topics.

Since topic modeling provides only an overarching summary of STEM teachers' discourse in the threads, we employed a two-stage manual review process. The qualitative coding process strengthens the validity of the machine-generated topics and better interprets how teachers' perceptions of GenAI are shaped not only by individual beliefs but also by peer discourse and emerging norms. Using the adjusted TAM framework (Kong et al., 2024), two authors independently coded 20% of the dataset, achieving substantial inter-rater reliability (Cohen's Kappa = .71). Following this, the first author coded the remaining dataset to ensure consistency. In total, 1248 discourse from the top 9 topics were coded.

Results

Topic modeling

Table 1 shows the top eight topics synthesized by our model, capturing 55.54% of the whole corpus. The first three predominant topics, "Using AI in Education", "Use of ChatGPT and "Lesson Plans and Essays" reflect the practical orientation of the teachers' online communities. User conversations are heavily focused on teaching practices, attitudes, and the integration of AI into daily educational routines. Discussions range from broad debates about the role of AI in classrooms to detailed accounts of teachers' experience with ChatGPT for lesson design and grading support. This underscores the increasingly pragmatic turn in teachers' grassroots discourse: not only how AI will impact education, but also how to adapt existing practices to incorporate or regulate it.

The next group of topics centers on academic integrity and student behaviors, highlighting anxieties over plagiarism and erosion of traditional teaching. Likewise, the "school policy" cluster shows educators' endeavor to detect plagiarism. Teachers raise concerns about fairness and responsibility, echoing prior literature on AI cheating and academic misconduct (Zhai & Krajcik, 2024).

Table 1
Top 8 Topics by Percentages

	Short Names	Percentage
1	Using AI in Education	18.24%
2	Use of ChatGPT	10.39%
3	Lesson Plans and Essays	9.29%
4	Students Cheating and Academic Integrity	5.86%
5	Student Education	5.64%
6	Kids Experience	2.25%
7	School Policy	1.98%
8	Scientific Calculators	1.89%

Content analysis

Across the coded dataset, Attitudes (ATT) is the most frequently identified dimension, representing 43.17% of the excerpts, followed by Perceived Usefulness (PU) at 18.43%, Self-Efficacy (SE) at 14.54%, and Subjective Norms (SN) at 10.86%. Behavioral Intention (BI) to use AI tools (7.43%) and Perceived Ease of Use (PEU) are the least at 5.57% (Table 2). These distributions suggest that STEM teachers' discussions are dominated by

affective perceptions of GenAI's usefulness, with relatively limited attention given to ease of use. Some selected excerpts are as follows.

Attitudes (ATT)

Teachers expressed more negative attitudes toward the use of AI in STEM education, with one of the most dominant concerns being students' ethical use of GenAI. As one participant reflected:

> Honestly, I'd be pretty happy if students could take an AI output, recognize whether it meets the criteria and technical requirements, fix the elements that are off. I mean, dang, they aren't even willing to do that with revisions on their own work. (ID 34)

This example illustrates that teachers are worried about students' overreliance on AI at the expense of creativity and problem-solving skills. While teachers expected authentic work, AI misuse has made originality harder to detect. As another teacher noted:

> It's super hard to detect AI-generated code; there's no signature or anything like that similar to image guardrails. Stylistically it could also be super inconsistent. (ID 64)

These concerns reflect a broader tension: the difficulty of detecting human-generated vs AI-generated content. To cope with this challenge, some teachers described proactive strategies, such as creating highly personalized rubrics, emphasizing process over product, handwritten notes on printed assignments, and increasing in-class presentations. Despite these concerns, a subset of teachers conveyed more positive attitudes, emphasizing the potential of AI if used responsibly. These teachers underscored the importance of preparing students for a future where AI is ubiquitous, while ensuring students still master foundational competencies.

Perceived usefulness (PU)

Teachers reported mixed experiences regarding the usefulness of generative AI in teaching. On the positive side, teachers acknowledged AI's potential to enhance their own instructional productivity, especially for time-consuming tasks, such as lesson planning, assignment generation, and preparing practice problems.

> SAVE ALL LESSON PLANS…every single thing you do! Some might suck and some might work but all are good ideas…until proven otherwise… and then… you just have to update them. I teach robotics and engineering…same process. (ID 758)

However, many described significant challenges when attempting to integrate AI into classroom practice. These challenges often reflected the limitations of the technology itself, such as "hallucinations," imprecise answers, or difficulties handling mathematical problems and formulas. As one teacher noted:

> It would have been great if it worked. I have been playing with AI for a while. I could not get AI to be precise enough. At best, it performs as a really bad, half-wit TA. (ID 33)

These comments highlight that teachers' judgments of usefulness are closely tied to reliability: when the tool produces errors, its perceived value in instructional contexts drops sharply. Teachers recognized the value of AI in streamlining routine tasks, while also cautioning that its outputs require adaptation, critical oversight, and professional judgment.

Self-efficacy (SE)

In terms of teachers' confidence in using AI, a dominant subtheme concerned the need to develop teachers' AI literacy. Several excerpts underscored the idea that teachers must first build their own competence before they can guide students in constructive use. For example, one participant emphasized,

> AI is not ruining education. Teachers need to learn how to use AI in the classroom. Teach your children how to use it in constructive ways. But first, you have to learn the same. (R39)

 © ISLS

Teachers with more positive attitudes toward generative AI frequently paired this optimism with explicit mentions of competency development. At times, teachers framed AI as a supportive student "peer", but emphasized that effective use depended on prompt engineering skills and responsible practices.

Subjective norms (SN)

Teachers' intentions to use AI were also shaped by contextual and social pressures. One participant reflected:

> Personal thoughts on general AI use in education aside, I think one of the problems with 'the push' to integrate it into classes and such is that a lot of the people loudly saying 'everyone should be using it, it's here and it's the future whether we like it or not, etc.' offer no actual ideas for how to use it. (R42)

This comment illustrates how peer pressure and broad societal discourse create pressure for adoption without sufficient guidance or concrete strategies. School policies also emerged as a key factor under the subjective norm category. Some teachers reported avoiding AI use because of institutional bans, while others expressed frustration at the lack of clear policy guidance from their universities/schools. This tension highlights how top-down directives, or their absence, shape teachers' willingness to experiment with AI. Finally, the teacher–student relationship appeared as a subtle but meaningful influence on adoption. As one participant explained:

> I get that there is a difference in using AI to do your homework and using AI to grade, but if I were a student and my teacher used AI when I wasn't allowed to, I would be at least a little bit upset. (R49)

Discussions

Most STEM teachers in this study expressed negative or skeptical attitudes toward the use of GenAI largely due to concerns about student ethics, the difficulty of detecting authenticity, and the lack of AI literacy. These findings align with prior research on challenges surrounding algorithmic bias, academic integrity, and the limitations of detecting AI-generated content (Zhai, 2025). However, they diverge from earlier accounts, suggesting that initial adopters of GPT tools tended to report more enthusiasm and optimism (Haque et al., 2022). This shift suggests that, as GenAI becomes more widely used in educational settings, teachers' perspectives are evolving toward a more cautious and critical stance.

Beyond documenting these attitudes, our findings reveal a key pattern that is less visible in prior work: teachers' acceptance of GenAI is role-dependent. STEM teachers in this study expressed greater willingness to use GenAI for their own professional tasks, such as lesson planning and content generation, while remaining more skeptical of its use by students. Rather than representing a uniform stance toward AI, teachers' perspectives vary across contexts of use, highlighting the need to conceptualize technology acceptance as situated and relational.

These findings also point to broader structural challenges shaping teachers' adoption of GenAI. Echoing Kumar and Sharma (2025) and Smith-Mutegi et al. (2025), teachers described tensions between expectations to adopt AI and the lack of clear guidance, policy, or institutional support. This highlights the role of subjective norms (SN) in shaping adoption, where external pressures may encourage use without providing sufficient pedagogical direction. Addressing this gap requires coordinated efforts across professional development, curriculum design, and policy frameworks.

This study contributes to learning sciences by examining large-scale, naturally occurring discourse data in online communities to understand how STEM teachers collectively make sense of GenAI use. Our findings clarify that teachers' acceptance of GenAI differs depending on the performance of tools, the AI literacy skills, and whether it's used for their own professional work or for students' learning. The findings suggest that effective integration of GenAI in STEM education requires moving beyond a generalized notion of "AI adoption" toward more differentiated approaches that explicitly address the practical strategies of AI use. Taken together, these insights highlight both the opportunities and tensions that GenAI introduces into STEM education. Teachers recognize AI's potential to reduce workload and enhance instructional efficiency, but their skepticism reflects unresolved challenges around trust, ethics, and pedagogy. Future research should prioritize co-designing AI integration strategies with teachers, ensuring that their expertise informs both the development of AI tools and the design of learning environments that support critical and ethical engagement.

References

Cheah, Y. H., & Kim, J. (2026). STEM teachers' perceptions, familiarity, and support needs for integrating generative artificial intelligence in K-12 education. *School Science and Mathematics*, *126*(1), 44–59.

Collie, R. J., & Martin, A. J. (2024). Teachers' motivation and engagement to harness generative AI for teaching and learning: The role of contextual, occupational, and background factors. *Computers and Education: Artificial Intelligence*, *6*, 100224.

Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, *13*(3), 319–340.

Egger, R., & Yu, J. (2022). A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts. *Frontiers in Sociology*, *7*, 886498.

Garofalo, S. G., & Farenga, S. J. (2025). Science Teacher Perceptions of the State of Knowledge and Education at the Advent of Generative Artificial Intelligence Popularity. *Science & Education*, *34*(2), 893–912.

Grootendorst, M. (2022). *BERTopic: Neural topic modeling with a class-based TF-IDF procedure* (arXiv:2203.05794). arXiv.

Haque, M. U., Dharmadasa, I., Sworna, Z. T., Rajapakse, R. N., & Ahmad, H. (2022). *"I think this is the most disruptive technology": Exploring Sentiments of ChatGPT Early Adopters using Twitter Data* (arXiv:2212.05856). arXiv.

Hsiao, C.-H., & Tang, K.-Y. (2025). Beyond acceptance: An empirical investigation of technological, ethical, social, and individual determinants of GenAI-supported learning in higher education. *Education and Information Technologies*, *30*(8), 10725–10750.

Kim, J. (2024). Leading teachers' perspective on teacher-AI collaboration in education. *Education and Information Technologies*, *29*(7), 8693–8724.

Kong, S. C., Yang, Y., & Hou, C. (2024). Examining teachers' behavioural intention of using generative artificial intelligence tools for teaching and learning based on the extended technology acceptance model. *Computers and Education: Artificial Intelligence*, *7*, 100328.

Kumar, R., & Sharma, S. (2025). Secondary school teachers' perspectives on GenAI proliferation: Generating advanced insights. *International Journal for Educational Integrity*, *21*(1), 7.

Ma, X., Hancock, J., & Naaman, M. (2016). Anonymity, Intimacy and Self-Disclosure in Social Media. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '16*, 3857–3869.

Salahshour Rad, M., Nilashi, M., & Mohamed Dahlan, H. (2018). Information technology adoption: A review of the literature and classification. *Universal Access in the Information Society*, *17*(2), 361–390.

Smith-Mutegi, D., Mamo, Y., Kim, J., Crompton, H., & McConnell, M. (2025). Perceptions of STEM education and artificial intelligence: A Twitter (X) sentiment analysis. *International Journal of STEM Education*, *12*(1), 9.

Venkatesh, V., & Davis, F. D. (2000). A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. *Management Science*, *46*(2), 186–204.

Wu, C., Wang, X., Carroll, J., & Rajtmajer, S. (2024). Reacting to Generative AI: Insights from Student and Faculty Discussions on Reddit. *Proceedings of the 16th ACM Web Science Conference, WEBSCI '24*, 103–113.

Zhai, X. (2025). Transforming Teachers' Roles and Agencies in the Era of Generative AI: Perceptions, Acceptance, Knowledge, and Practices. *Journal of Science Education and Technology*, *34*(6), 1323–1333.

Zhai, X., & Krajcik, J. (Eds.). (2024). *Uses of artificial intelligence in STEM education*. Oxford University Press.

International Society of the Learning Sciences

Teacher Perception and Practices of the Next Generation Science Standards (NGSS): Insights from a New England Study

Anupam Raj, Columbia University, ar5099@tc.columbia.edu
Michael C. Wittmann, American Physical Society, wittmann@aps.org

Abstract: This qualitative study explores how secondary science teachers understand and implement two core dimensions of the Next Generation Science Standards (NGSS): Science and Engineering Practices (SEPs) and Crosscutting Concepts (CCCs). Drawing on semi-structured interviews with ten teachers from schools serving economically diverse communities in a New England state, the analysis reveals distinct orientations toward SEPs. These orientations range from identity-based and assessment-aligned approaches to emerging understanding through professional learning. In contrast, CCCs appeared less intentionally addressed, with most teachers using them as implicit outcomes rather than explicit instructional goals. A few teachers, however, used CCCs to promote interdisciplinary connections and support comparative reasoning. These findings highlight variation in teachers' enactment of NGSS dimensions and the need for ongoing professional learning that supports integration of SEPs and CCCs. The study contributes to efforts to bridge the gap between policy vision and classroom realities in three-dimensional science teaching and assessment.

Introduction

Science education in the United States is changing. In 2013, states adopted the Next Generation Science Standards (NGSS) to align classroom learning with a new national vision (NGSS Lead States, 2013). That vision comes from *the Framework for K–12 Science Education,* which calls for three-dimensional learning that blends disciplinary core ideas, science and engineering practices, and crosscutting concepts (National Research Council [NRC], 2012). The goal is coherence. Students should learn science by doing science and by seeing ideas that connect across topics (NRC, 2012; NGSS Lead States, 2013). Professional groups echo this shift and describe three-dimensional teaching and learning as the expected baseline for modern science instruction (National Science Teaching Association, 2018). Research has shown that when students engage in Science and Engineering Practices (SEPs) and Crosscutting Concepts (CCCs) alongside disciplinary ideas, they develop deeper conceptual understanding, stronger reasoning skills, and greater motivation to learn science (He et al., 2024; Nordine & Lee, 2023; Sherwood & Benedict-Chambers, 2025).

Yet a decade on, research continues to show that the implementation of the NGSS vision remains uneven. Classrooms vary widely in how deeply teachers integrate the three dimensions or how consistently they connect them in instruction (He et al., 2024). Studies reveal that while teachers often adopt the language of NGSS, they struggle to coordinate the disciplinary ideas, practices, and crosscutting concepts in meaningful ways (Nordine & Lee, 2023; NRC, 2012). A recent analysis showed that implementation of three-dimensional science in secondary schools remains lower than expected nationwide (Smith, 2020) and may be influenced by the complex interplay of factors such as access to professional development, curriculum resources, teacher knowledge, and administrative support (Haag & Megowan-Romanowicz, 2021; He et al., 2024). Together, these studies show that three-dimensional learning often remains *aspirational*, shaped by contextual constraints, uneven professional learning, and differing interpretations of what it means to teach science as practice.

Within this uneven landscape, Science and Engineering Practices have received the most attention. Teachers commonly report that engaging students in SEPs helps make learning active and authentic (He et al., 2024; Krajcik & Merritt, 2012). While studies have demonstrated the promise of SEPs for improving student engagement and understanding (Kang et al., 2019), research also highlights persistent challenges in enacting them effectively. In a large-scale study across more than 1,200 schools, Banilower et al. (2018) found that science teachers at both elementary and secondary levels struggled to integrate SEPs meaningfully into instruction. These findings suggest that while SEPs are central to the NGSS vision, consistent implementation remains difficult.

Crosscutting Concepts, on the other hand, are the least visible dimension of NGSS in classroom instruction (Nordine & Lee, 2023; Sherwood & Benedict-Chambers, 2025). Although CCCs are intended to

 © ISLS

promote coherence and transfer across topics, some evidence shows that teachers rarely mention or assess them directly (Nordine & Lee, 2023). Even when teachers acknowledge CCCs conceptually, few report using them to guide instruction or assessment design (He et al., 2024). This limited engagement suggests that CCCs, while valued in theory, remain difficult to translate into practice.

Thus, understanding how teachers interpret, practice, and assess both SEPs and CCCs is critical for explaining the persistent gap between the NGSS framework and its classroom realities. Building on this context, this study explores *how secondary science teachers understand and enact the three-dimensional vision of the NGSS in their daily instruction*. Mainly, it focuses on how teachers interpret and use Science and Engineering Practices (SEPs) and Crosscutting Concepts (CCCs) in teaching and assessment.

NGSS framework

The present study is grounded in the three-dimensional learning model proposed by *the Framework for K–12 Science Education* (NRC, 2012) and operationalized through the Next Generation Science Standards (NGSS Lead States, 2013). The NGSS defines science learning as the integration of disciplinary knowledge (via DCIs), epistemic practices (via SEPs), and conceptual connections (via CCCs) across scientific domains. This framework reconceptualizes science education as active sensemaking—where students explain, model, and reason about phenomena using the tools and habits of scientists (Krajcik & Merritt, 2012; Penuel & Reiser, 2018). Rather than treating the dimensions as separate strands, the framework emphasizes their interaction: SEPs provide ways of knowing, CCCs connect ideas across contexts, and DCIs anchor understanding in key disciplinary ideas. Together, they aim to develop students who can apply scientific reasoning flexibly and coherently (Laverty et al., 2016; NRC, 2012).

Using NGSS as a theoretical lens allows this study to analyze how teachers conceptualize and enact these intertwined dimensions within real classroom systems. Prior work has shown that the success of three-dimensional instruction depends heavily on teachers' *interpretations* and *contextual* constraints (He et al., 2024; Lee & Arias, 2024). Teachers' understanding of SEPs and CCCs, in particular, serves as a window into how the NGSS vision is translated into daily teaching and assessment. Framing the study through the NGSS therefore provides both a conceptual map and an evaluative lens for examining how teachers make sense of and implement the two key dimensions of three-dimensional learning.

Methodology

This qualitative study uses an interpretivist approach (Creswell & Poth, 2018; Schwandt, 2000) to explore secondary science teachers' understanding and enactment of the three-dimensional vision of the Next Generation Science Standards (NGSS), focusing on SEPs and CCCs. The research draws on data originally collected to examine how secondary science teachers adapted instruction during the COVID-19 lockdown in Spring 2020 (Raj, 2022). Subsequent analyses of this dataset have examined equity (Raj et al., 2024), differentiated instruction (Raj et al., 2025a), and teacher resilience (Raj et al., 2025b). Building on this work, the present study draws on the same interviews to explore teachers' reflections on NGSS-aligned practices, offering insight into how they conceptualize and implement SEPs and CCCs.

Participants included ten secondary science teachers (six male, four female) from a New England state, representing high-, middle-, and low-socioeconomic contexts. Teachers were recruited using a combination of convenience and purposeful sampling. Semi-structured Zoom interviews were conducted during the summer of 2020, lasting between 45 minutes and two hours. Conversations focused on teachers' instructional adaptations, assessment approaches, and perspectives on the three dimensions of science learning. Although participants taught in varied socioeconomic settings, the present analysis did not reveal systematic differences, so this paper centers on shared conceptions and practices.

Data were analyzed using Braun and Clarke's (2006) thematic analysis framework. The process began with inductive coding to identify salient ideas and patterns, followed by deductive refinement guided by the NGSS dimensions—particularly SEPs and CCCs—which served as analytic lenses for interpretation. This iterative approach allowed for an integrated understanding of teacher sensemaking, capturing both emergent and theory-informed insights. The findings below highlight how teachers in our study interpret and apply SEPs and CCCs within their classroom practice.

Findings

Spectrum of science and engineering practices

While most teachers mentioned that Disciplinary Core Ideas (DCIs) were relatively easier to teach and assess, their engagement with Science and Engineering Practices (SEPs) ranged from full integration to general

awareness. For some teachers, SEPs are not simply pedagogical strategies but an *identity* to cultivate. Hank, for instance, viewed his classroom as a space where students gradually *become* scientists. He explained, "*I tell my students that I do expect them to be functioning physicists at the end of the year… I spend a lot of time focusing on defining questions, creating experiments to answer those questions…creating models… and the critical thinking skills involved in assessing those models.*" His expectation that students "be functioning physicists" reveals how he treats SEPs as habits of mind rather than an isolated academic skill. Similarly, Heather echoed this framing when she explained that students "*ask questions…design a way to look for answers…we use CER method where they make a Claim…give Evidence…and then the Reasoning behind that.*" Her emphasis on the CER method reflects an effort to help students practice structured scientific reasoning. Together, these examples illustrate teachers' mindset and practices when integrating SEPs into their instruction.

At the same time, several teachers approached SEPs through structured systems of assessment and accountability. For instance, Liam explained, "*All of my grades are assigned to a science and engineering practice…when I make an assignment…I'm a little more intentional about what types of questions those assessments have, because I want to make sure that I can really justify that it does apply very clearly to one of those practices.*" His reflection shows a deliberate effort to align every task with a specific practice, ensuring clarity between teaching and evaluation of SEPs. Similarly, Mike described how his school's graduation system formally embeds SEPs into accountability structures: "*We just assigned the science skills to [school name] high school graduation standards. So, students have to show proficiency in those skills to graduate.*" This example highlights a school-level commitment where SEPs are not only taught but also institutionalized as criteria for demonstrating competence. Together, these accounts illustrate classrooms and school systems that have established measurable and transparent indicators of SEPs-related learning.

While some teachers demonstrated confidence in integrating SEPs, others were still developing a deeper understanding through professional learning. Mia shared, *"I've come a long way with it, because I've had lots of opportunities with professional development…a lot of this is that discovery piece and allowing students to discover a lot of these things on their own, and then applying it in that assessment."* Her reflection highlights how sustained professional learning can strengthen teachers' confidence and shape their teaching practices with SEPs.

By contrast, Lydia described navigating structural barriers that limited her ability to fully implement SEPs. She explained, *"I'm just getting, I'm just delving into this [NGSS]… I wanted our power standards to be the practices… but that was not what was desired by administration… my assessments are going to be tied to the core content…without looking specifically at the crosscutting concepts or SEPs…I do post them all in my classroom… I do let students know what the crosscutting concepts are… and I do bring them up when I introduce the course content standards."* Her comments reveal both awareness and limitation as she values SEPs but works within a system that prioritizes content coverage over integrated practices.

Across teachers, these differences reflected three orientations: identity-driven, system-driven, and evolving through professional learning. In contrast, some teachers, like Lydia, faced institutional constraints that limited their ability to move beyond content-focused instruction. Together, these perspectives show that teachers' engagement with SEPs varies widely, shaped by their experiences and school contexts.

Limited Crosscutting Concepts in practice

Crosscutting Concepts (CCCs) were the least discussed and least intentionally assessed dimension of NGSS among our participants. Most teachers did not even bring up CCCs when reflecting on their approach to NGSS during remote instruction. They focused instead on how they handled core content and attempted different ways to include SEPs in their teaching and assessment. Considering that CCCs were not the direct topic of investigation, we did not probe deeply during interviews. However, some evidence suggests that teachers' attention to CCCs was minimal and often implicit.

For instance, Heather admitted, *"Yeah, I don't think I have that intentionally assessed… Let me think. Yeah… I definitely didn't consciously think of ways to assess that [CCCs] differently."* Her hesitation captures the uncertainty many teachers felt about where CCCs fit within their instruction. Heather's reflection reveals that she does not reject the idea of CCCs but had not yet found clear ways to integrate or assess them within her existing routines. This suggests that while teachers were aware of CCCs conceptually, they were still exploring how to connect them meaningfully to instruction and assessment.

Hank, who otherwise demonstrated a highly intentional approach to SEPs above, expressed a similar stance. He reflected, *"I do not focus much on the crosscutting concepts before or after remote instruction… I use the content standards to inform me and then make sure they apply and are clearly applicable to SEPs."* His statement shows that even among teachers with well-developed assessment systems for SEPs, CCCs often fall outside their instructional priorities. Both Heather and Hank's reflections suggest that CCCs are acknowledged but not always translated into specific instructional or assessment practices. While most participants could name

the seven crosscutting concepts—patterns, cause and effect, scale, systems, energy and matter, structure and function, and stability and change—few reported weaving them systematically into lessons.

We did find two disconfirming examples where teachers described using CCCs in more intentional or natural ways. Hazel explained, *"crosscutting concepts is not really a big problem to assess remotely, because it can often be done with a writing piece or being able to look at similarities and differences… my friends in environmental science designed their entire end of the year around looking at these kinds of crosscutting concepts."* Her comment reveals a clear awareness of CCCs as a thinking framework that encourages comparison, and connection. Similarly, another teacher, Luke, described a more interdisciplinary perspective when talking about CCCs. He explained, *"We did with our freshman group… we have two interdisciplinary units that we do with Social Studies and English. One of them… was about the changing nature of our community… from the science end of it, we look at human impact… of our activities on our communities."* Luke's reflection shows how CCCs can connect ideas across subjects and act as an opportunity for inter-subject collaboration among teachers, allowing students to see how scientific and social perspectives intersect.

Overall, teachers' engagement with crosscutting concepts mirrored what we observed with Science and Engineering Practices, showing a spectrum of understanding shaped by context and support. Some teachers interpreted CCCs as implicit patterns that emerge during inquiry, while others actively integrated them through writing or interdisciplinary projects. However, for many, CCCs remained less defined and less intentionally embedded compared to SEPs. These variations suggest that while both dimensions share the potential to deepen reasoning and connection making, CCCs are often the least implemented part of three-dimensional learning.

Discussion and implications

For SEPs

Findings from this study highlight a clear range in how teachers understood and used SEPs. Some, like Hank and Heather, viewed SEPs as identity-building tools, encouraging students to see themselves as scientists through activities such as modeling, designing experiments, and using evidence-based reasoning. Others, such as Liam and Mike, approached SEPs through structured assessment systems that linked grades and graduation standards directly to specific practices. A few, like Mia, described themselves as still learning to use SEPs intentionally through professional development. In contrast, Lydia highlighted the institutional barriers that limited this process, noting that her school's emphasis on core content often left little room to prioritize SEPs. These variation mirrors earlier research showing that while teachers value SEPs for promoting active and meaningful science learning, their classroom use often depends on experience, training, and context (He et al., 2024; Krajcik & Merritt, 2012).

These results reinforce the need for systemic supports that help teachers engage meaningfully with SEPs. As He et al. (2024) found, strong implementation depends on access to professional learning, collaboration, and reform-oriented instructional resources. Our findings similarly suggest that even motivated teachers struggle to sustain SEP-focused instruction without institutional alignment and shared structures for growth. Educational institutions must prioritize comprehensive support systems that include rich materials, institutional support, and sustained professional development (Brand, 2020; Luft et al., 2022).

For CCCs

Teachers in this study seldom described teaching or assessing crosscutting concepts (CCCs) intentionally, a finding consistent with research trends (Nordine & Lee, 2023). While most participants were familiar with the CCCs, many found them difficult to apply in everyday instruction. Teachers such as Hank acknowledged that they did not focus on CCCs directly, often prioritizing content coverage or SEPs. This pattern reflects earlier work showing that CCCs are often treated as implicit outcomes rather than planned goals (Fick & Arias, 2022).

At the same time, several teachers showed that CCCs can be integrated naturally when supported by written assignments or collaboration. Hazel described using writing and comparison tasks to help students recognize relationships and patterns, while Luke shared interdisciplinary lessons linking environmental science with social studies and English. Their approaches align with Nordine and Lee's (2023) description of CCCs as tools for connecting ideas across contexts and supporting sensemaking. Lee and Arias (2024) similarly argue that implementing CCCs requires *complex* and *multifaceted* knowledge, combining content understanding with pedagogical skill. Fick and Arias (2022) note, thoughtful use of CCCs can also promote equity by making science learning more connected and accessible. Taken together, these insights suggest that teachers are aware of CCCs' potential but need sustained professional learning to help them apply these concepts intentionally in classroom.

References

Banilower, E. R., Smith, P. S., Malzahn, K. A., Plumley, C. L., Gordon, E. M., & Hayes, M. L. (2018). *Report of the 2018 NSSME+ (National Survey of Science and Mathematics Education).* Horizon Research, Inc.

Brand, B. R. (2020). Supporting elementary teacher learning through science education reform: Science learning and leadership for elementary school teachers (SLL). *Research in Science Education, 50*(2), 441–463.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches (4th ed.)*. SAGE Publications.

Fick, S. J., & Arias, A. M. (2022). A review of literature that uses the lens of the next generation science crosscutting concepts: 2012–2020. *Journal of Research in Science Teaching, 59*(6), 884–913.

Haag, S., & Megowan-Romanowicz, C. (2021). Building competence in science and engineering. *Standards, 1*(1), 39–52. https://doi.org/10.3390/standards1010005

He, Y., Park, M., & Blank, R. (2024). Ten years of three-dimensional science and its implementation in the secondary classroom: A scoping review. *Science Education Review, 112*(4), 1–25.

Kang, E. J., McCarthy, M. J., & Donovan, C. (2019). Elementary teachers' enactment of the NGSS science and engineering practices. *Journal of Science Teacher Education, 30*(7), 788–814.

Krajcik, J., & Merritt, J. (2012). Engaging students in scientific practices: What does constructing and revising models look like in the science classroom? *The Science Teacher, 79*(3), 38–41.

Laverty, J. T., Underwood, S. M., Matz, R. L., Posey, L. A., Carmel, J. H., Caballero, M. D., Fata-Hartley, C. L., Ebert-May, D., Jardeleza, S. E., & Cooper, M. M. (2016). Characterizing college science assessments: The three-dimensional learning assessment protocol. *PLOS ONE, 11*(9), e0162333.

Lee, O., & Arias, A. M. (2024). Elementary preservice teachers' initial knowledge for teaching related to crosscutting concepts within three-dimensional learning and teaching. *Science and Mathematics Education, 115*(1), 1–22. https://doi-org.tc.idm.oclc.org/10.1080/0047231X.2024.2390297

Luft, J. A., Wong, S. S., & Zacharia, Z. C. (2022). Professional development for science teachers: Current research and future directions. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of Research on Science Education* (Vol. 3, pp. 401–425). Routledge.

National Research Council. (2012). *A framework for K–12 science education: Practices, crosscutting concepts, and core ideas.* The National Academies Press. https://doi.org/10.17226/13165

National Science Teaching Association. (2018). *NSTA position statement: The Next Generation Science Standards.* https://www.nsta.org

NGSS Lead States. (2013). *Next Generation Science Standards: For states, by states.* The National Academies Press.

Nordine, J. C., & Lee, O. (2023). On the Nature and Utility of Crosscutting Concepts. *Education Sciences, 13*(7), 640. https://doi.org/10.3390/educsci13070640

Penuel, W. R., & Reiser, B. J. (2018). Designing NGSS-aligned three-dimensional assessments. *K-12 STEM Education, 4*(3), 25–36.

Raj, A. (2022). *Investigating the teaching and assessment experiences of Maine secondary science teachers during the COVID-19 lockdown*. Electronic Theses and Dissertations. 3633.
https://digitalcommons.library.umaine.edu/etd/3633

Raj, A., Wittmann, M., & Kayumova, S. (2024). Being fair & equitable: A qualitative study of science teachers' shifts in priorities and expectations during COVID-19. In *Proceedings of the 18th International Conference of the Learning Sciences (ICLS)*. International Society of the Learning Sciences. https://repository.isls.org/bitstream/1/10977/1/ICLS2024_2283-2284.pdf

Raj, A., Wittmann, M., & Kayumova, S. (2025a). Rethinking differentiated instruction: Lessons from the pandemic for today's diverse classrooms. In *Proceedings of the 19th International Conference of the Learning Sciences – ISLS 2025*(pp. 3200–3203). International Society of the Learning Sciences.

Raj, A., Wittmann, M., & Kayumova, S. (2025b). Resilient and transformative journey of teachers through the pandemic and beyond. In *Proceedings of the 19th International Conference of the Learning Sciences – ISLS 2025* (pp. 2115–2119). International Society of the Learning Sciences.

Schwandt, T. A. (2000). Three epistemological stances for qualitative inquiry: Interpretivism, hermeneutics, and social constructionism. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 189–213). SAGE Publications.

Sherwood, C.-A., & Benedict-Chambers, A. (2025). Elementary preservice teachers' use of the crosscutting concepts to plan NGSS-aligned lessons. *Science Education, 109*(2), 1–15.

Smith, P. S. (2020). What does a national survey tell us about progress toward the vision of the NGSS? Journal of Science Teacher Education, 31(6), 601–609. https://doi.org/10.1080/1046560X.2020.1786261

Symposia

 © ISLS

 © ISLS

International Society of the Learning Sciences

Designing AI Partner-Embedded Learning Environments to Support AI Literacies

Mon-Lin Monica Ko (Chair), University of Colorado Boulder, monlin.ko@colorado.edu
Marcelo Worsley (Discussant), Northwestern University, marcelo.worsley@northwestern.edu
Michael Alan Chang, Boston University, machang@bu.edu

Chelsea Chandler, Sierra Rose, Brooklyn Clines, Emily Watts, John Weatherley, Peter W. Foltz, Rohit Raju, Roshan Muddaluru, Jeffrey Bush Tayne, Greg Benedis Grab, Quentin Biddy, Thomas Breideband, William R. Penuel, Tamara Sumner, Collette Heskett, Nga Hoang, Julia Eden, Sidney D'Mello
Chelsea.Chandler@colorado.edu, Sierra.Rose@colorado.edu, Brooklyn.Clines@colorado.edu, Emily.Watts@colorado.edu, John.Weatherley@colorado.edu, peter.foltz@colorado.edu, rohit.raju@colorado.edu, Roshan.Muddaluru@colorado.edu, jeff.bush.tayne@Colorado.EDU, Greg.Benedis-Grab@colorado.edu, Quentin.Biddy@Colorado.EDU, Thomas.Breideband@colorado.edu, William.Penuel@colorado.edu, tamara.sumner@colorado.edu, Collette.Heskett@colorado.edu, Nga.Hoang-1@colorado.edu, Julia.Eden@colorado.edu, Sidney.Dmello@Colorado.EDU
University of Colorado Boulder

Abstract: As AI technologies become increasingly integrated into schools, questions arise about what students should learn about AI and how AI-powered tools can support collaborative learning. This symposium reports on the Institute of Student-AI Teaming (iSAT), which designs AI partners and curricular activity systems to promote AI literacy and collaboration. Unlike most classroom AI tools designed for 1:1 interactions, iSAT develops AI partners as social collaborators that both facilitate group learning and serve as objects of inquiry—helping students "look under the hood" of AI systems. Our multi-year co-design process across three U.S. states engages teachers, students, and parents to build integrated systems of curriculum, technology, and professional learning. The symposium's six posters span AI design, infrastructure, curriculum, and teacher learning, illustrating how AI can be productively embedded within classroom collaboration. Together, they advance understanding of how AI systems can meaningfully support and transform collaborative learning.

Objective of the session

As AI technologies become increasingly embedded in education (Reich & Dukes, 2024), so do questions about what students should learn about AI systems (AI literacy) and how AI-powered tools may support (or hinder) student learning. While there is growing excitement about AI's potential to transform education, decades of research highlight repeated failures of earlier innovations to do so. Scholars argue these failures often stem from *underappreciating* the infrastructural nature of technology and *overestimating* its power to drive change in classroom practice (Roschelle et al., 2010). First, technology is infrastructural in that its value lies not in its particular content but in its ability to generate new representations of disciplinary ideas and create opportunities for students to engage more deeply in those ideas (Hegedus & Kaput, 2003). Second, like previous technologies, today's AI tools will not transform education on their own. Their success depends on how curriculum activity systems—teacher professional learning (PL), instructional materials, and technology—work together to reshape learning environments (Vahey et al., 2013). Their success also depends on how teachers and students work together within the system to support learning.

This symposium reports on work from the Institute of Student-AI Teaming (iSAT), an institute that aims to design AI-powered tools and curricular activity systems that promote collaborative knowledge building and student learning around AI-literacy specifically (D'Mello et al., 2024, OECD, 2025). Our project intentionally combines both AI-literacy *and* AI as a tool for collaborative learning in ways that will enable the AI-partner to serve simultaneously as a collaboration support and an object of inquiry as students think critically about how the AI works. That is, our AI partners play a dual role of supporting student collaboration and as a tool for "looking under the hood" at how AI systems work. Supporting collaborative learning is a longstanding goal of both the ISLS and CSCL community, and is crucial because we know that learning is inherently a social process. Yet, the dominant approach to incorporating AI inside K-12 classrooms focuses on 1:1 interactions between students and AI systems. iSAT's approach is radically different: we aim to develop AI as a social, collaborative partner that works alongside students and teachers to deepen their engagement with one another. This goal has required that our team engage in multi-year co-design processes to better understand the needs and desires of multiple users

and stakeholders (e.g. teachers, students, parents) as we work to develop and refine AI partners, curriculum materials, and professional development to support the development of AI literacies. Over the past five years, our work has involved working with stakeholders in districts across three U.S. states to conceptualize, develop, enact, and refine our curricular activity system.

This session will focus on the design of iSAT's curricular activity system to deepen the field's conversation about the role of AI-systems in supporting collaborative learning. Across the posters, we will describe how our co-design practices, back-end models, classroom pilot studies, professional learning, and curricular routines and designs worked toward our emerging understanding of the utility of these tools for supporting collaborative learning.

Overview of the symposium

The 6 posters that constitute the symposium represent our collective efforts to take seriously what it means for AI systems to *partner* within the curricular activity systems of secondary schools. Poster 1 will detail the trajectory of co-design for our first AI partner, the Community Builder (CoBi) and how the ongoing design, implementation and analysis of the co-design space informed the values of our institute. Poster 2 will present on the design, development, and pilot studies of our second Jigsaw Interactive Agent (JIA) in secondary classrooms and how productive uncertainty was leveraged to design the AI partner interventions to support small group jigsaws. In Poster 3, we will report on the architecture and analytic pipeline that underlies our two existing AI partners, documenting how the pipeline infrastructure, state detection and dialogue policies inform the AI-partner outputs and interventions. Posters 4-6 address how we built both curriculum materials and teachers' capacities to leverage them so that the AI partners could be meaningfully integrated into everyday classroom activities. In Poster 4, we examine how curricular routines can support students' sensemaking of their own collaborative practices. In Poster 5, we show how our PL design supported teachers' learning about AI, our AI literacy curriculum and pedagogical approaches that center student sensemaking. Finally, Poster 6 describes the design and refinement of our curriculum materials that focus on AI literacy, a process guided by co-design, adaptation and refinement with teachers in their classrooms.

Scholarly significance

This session is both timely and essential. First, AI tools are becoming increasingly embedded in schools and in our world. At the same time, few teachers are currently prepared to teach about AI, use AI-enabled tools, and adopt student-centered, collaborative pedagogies—what we call a *triple innovation challenge*. We also know that external forces (e.g. social and news media) shape students' understanding of AI, and that schools/educators play a less significant role (OECD, 2025) This session highlights a new approach to supporting AI literacy through the design of curricular activity systems where AI Partners are meaningfully embedded to support AI literacies through collaborative learning.

Structure of the session

The symposium will be organized as a structured poster session. The chair of the session will first provide a brief overview of the session (4 mins). Each presenter will give a one-minute elevator pitch to introduce their project (6 mins). Presenters will then move to their respective posters; participants can view posters, demos (if applicable) and engage in discussion. We will invite participants to engage in two "rounds" of poster-viewing (30 mins each), so that co-presenters have a chance to view others' posters . Our discussant will synthesize the posters and pose questions for the audience to discuss, focusing on the challenges and affordances of AI-powered systems in supporting collaborative learning in schools (20 minutes).

CoBi Origins: Co-design narratives as boundary objects to translate Co-design findings into interdisciplinary implementation

Michael Alan Chang

Objectives

Collaborative and participatory design is increasingly used to inform the design and implementation of AI-Ed tools that are responsive to the needs and lived experiences of interest holders (Ravi et al., 2025). An AI-Ed research team hosts a co-design context, *completes* an analysis of it, and *then* responds to the co-design efforts through implementation. While this sequential and linear process simplifies the already complex and interdisciplinary process of building AI-Ed tools, it can constrain both the types of analysis that can be conducted within the co-design context and the ways in which stakeholders can engage in the ongoing design of AI-Ed tools

(Chang, *under review*). Here, I share our efforts in iSAT, where there were parallel, longitudinal efforts to (a) develop and analyze a three-year co-design study with high school youth to surface their hopes and concerns around the use of AI in schools and (b) respond to the findings of those co-design studies to develop AI partners and curricular activity systems that surround the use of the technology. Here, I ask: how does the ongoing design, implementation, and analysis of co-design spaces shape the design activities of an interdisciplinary research institute?

Perspectives

Here, I view *stories about co-design activities* as an evolving boundary object that guides how large-scale AI-Ed implementation teams take on co-design findings (Star & Griesemer, 1989). Boundary objects are conceptualized as "how artifacts can fulfill a specific function in bridging intersecting practices" (pg. 134, Akkerman & Bakker, 2011). As Star (2010) clarifies, a boundary object is characterized by interpretive flexibility across different contexts, allowing each context to tailor the object to its own needs locally. Finally, boundary objects follow particular trajectories within organizational contexts; once created, boundary objects evolve through local tailoring, standardize over time, and, finally, fade away as the standardization practices begin to constrain possible local adaptations. In turn, new boundary objects are then created.

Methods

Along with several other iSAT members, I co-led the development of three co-design workshops that engaged youth of color in envisioning expansive possibilities for AI-supported collaboration (Chang et al., 2024). During that time period, iSAT also collectively began to explore the findings from the workshops. Here, I study the presentations I (and also other iSAT members) created to share these findings, both internally and externally. Additionally, I examine the field notes from discussions held during design meetings to build the Community Builder (CoBi), a tool created by iSAT that was directly informed by prototypes developed by youth participants during our co-design sessions (Breideband et al., 2023). Across these data sources, I qualitatively coded for instances where meeting participants referenced the narrative about how CoBi was created.

Results

Initially, the CoBi co-design narrative described how youth participants in our co-design engagements visited a graduate student cooperative house, where they learned how members of that cooperative house collectively created agreements about house policies, which they were then held accountable for. This cooperative house then inspired a new AI Tool, CoBi, that supported collaboration by helping youth stay accountable to collaborative norms that they *themselves* deemed to be important. Institute members repeatedly described how this narrative helped them understand the collective commitments and pedagogical values embodied in the youth's envisionings of CoBi, beyond what could be conveyed simply through the youth participants' drawings of CoBi. Over time, this narrative about co-design became "standardized" through both internal and external facing presentations, as slides about this narrative were re-purposed and shortened. Concurrently, my own deeper analysis of the co-design data introduced new complexities about CoBi. For instance, CoBi had not only sprung from the cooperative house visit but also from the participants' developing political identities; indeed, CoBi was exciting to the youth primarily because it could be envisioned to support their collective agency in creating better school-wide conditions.

Significance

Large-scale interdisciplinary teams building AI-Ed tools now often make commitments to responsible innovation (Stilgoe et al., 2013). This work demonstrates how boundary objects from co-design spaces can both accelerate interdisciplinary collaboration and responsiveness to the co-design space, while simultaneously constraining the uptake of particular ideas, particularly as analysis of co-design efforts deepens and shifts. This work points to the need for constant iteration and collaboration in the use and revision of boundary objects.

Poster 2: Jigsaw Interactive Agent (JIA): Pilot Study of JIA's Real-time Support for Collaboration through Productive Uncertainty Interventions

Mon-Lin Monica Ko, Chelsea Chandler, Sierra Rose, Brooklyn Clines, Emily Watts, John Weatherley, Peter W. Foltz

Objectives

This study investigates how an AI partner can support small-group collaboration in a middle school classroom jigsaw activity. While jigsaws have long been used to structure collaborative learning, their success depends on high-quality facilitation. Teachers face challenges in providing real-time, tailored support across multiple groups of students. The preliminary designs of our AI partner were tested in the lab (see Doherty et al., 2025). We designed an AI partner to generate just-in-time interventions that foster productive collaboration, aiming to extend CSCL research on how artificial intelligence can scaffold multi-party learning interactions.

Perspectives

Drawing on existing work on collaborative problem solving (Sun et al., 2020), we position AI not as a replacement for teachers, but as a partner that captures multiparty discourse and generates timely prompts to support social, cognitive, and metacognitive processes in group learning. Drawing on scholarship on productive uncertainty (Manz & Suarez, 2018), we designed the partner to intervene in 4 ways: provide social support, connect ideas, problematize assumptions, and stabilize group progress. This study investigates both the accuracy and appropriateness of the AI partner interventions as well as students' responses to them.

Methods

We embedded the AI partner within a single lesson from a co-designed AI literacy curriculum implemented in two school districts with students aged 13-14 years old (N=64 groups). In this lesson, students engaged in a jigsaw task where they read about different moderation approaches, generated criteria for evaluating these approaches, and then worked collaboratively to summarize the approaches and rank them based on their ability to foster inclusive online communities. In the context of this lesson, the AI partner, the Jigsaw Interactive Agent (JIA) analyzed student talk in real time using a multimodal pipeline and provided one-minute interval interventions via a classroom web interface. We tested three approaches to JIA (rule-based, LLM-based, hybrid). Data sources included audio, video, log files, and 347 AI–student interaction transcripts. Human annotators evaluated (a) appropriateness of interventions, (b) student uptake, and (c) emotional responses.

Results

The LLM and hybrid models generated contextually appropriate interventions 82–87% of the time, compared to 63% for the rule-based model. Most interventions focused on connecting (50%) and social support (26%). Although 61% of the AI partner interventions received no reaction from students, when students *did* notice the intervention, 28% of students' responses integrated the AI partner's feedback into group reasoning, deepening or shifting the trajectory of their subsequent discussion. Emotional responses were mostly neutral or positive, with 78% of interactions judged as authentic through human annotations. Instances of genuine uptake included students' increased attention to criteria when ranking the approaches, weighing tradeoffs between the approaches, and incorporating aspects of the AI partner intervention in their subsequent conversations.

Significance

This study demonstrates how AI partners might be able to extend teacher facilitation in CSCL contexts by supporting multiparty discourse and fostering productive uncertainty. The findings highlight design considerations for AI partners: the importance of integrating rule-based collaboration analytics with adaptive LLM feedback, calibrating the timing and framing of interventions, and exploring new embodiments to increase student uptake. This work also underscores that evaluating AI in collaborative learning requires attending to how feedback is noticed, integrated, and trusted by students. These insights advance the design of AI systems that authentically scaffold collaboration in real-world classrooms.

Poster 3: AI-based methods to assess and support small group collaboration through discourse

Chelsea Chandler, Rohit Raju, Roshan Muddaluru, Mon-Lin Monica Ko, John Weatherley, Sidney D'Mello, Peter W. Foltz

Objectives

Collaborative discourse in classrooms represents a rich source of data for understanding and supporting learning. Yet despite advances in discourse analytics, current AI tools often fall short in interpreting the full context of student dialogue, or aligning feedback to pedagogically meaningful constructs (Graesser et al., 2018). Our objective is to design and evaluate AI partners that can analyze and respond to student collaboration in principled

 © ISLS

ways anchored in established frameworks of collaboration and pedagogical practice. By grounding large language models (LLMs) in frameworks such as Productive Uncertainty (Manz & Suarez, 2018), Collaborative Problem Solving (CPS) (e.g., Sun et al., 2020), and Community Agreements (e.g., OpenSciEd), we aim to create AI partners capable of both assessing and scaffolding higher-order collaborative and metacognitive skills in real time.

Perspectives

Prior research has shown that "traditional" Natural Language Processing (NLP) approaches - such as discriminative models trained to detect discourse moves or collaborative indicators - can achieve reasonable accuracy (e.g. Dowell et al., 2019). However, such models are typically limited to surface-level features, local utterances, and limited context. They struggle to capture the dynamic, multi-turn, and socially situated nature of collaborative learning. In contrast, LLMs offer the potential for deeper contextual understanding and generative, real-time feedback, enabling more adaptive and pedagogically aligned forms of support. The challenge lies in ensuring that these models' responses are not just accurate in assessing collaborative states and fluent in their feedback, but also that they are instructionally principled. This work brings a dual perspective: (1) advancing AI's capacity to understand collaborative learning as a social-cognitive process, and (2) ensuring that AI-generated interventions align with established theories of learning and assessment constructs rather than data-driven heuristics.

Methods

We describe our experience in the design and evaluation of two distinct AI partners: CoBi (the Community Builder, based on CPS-mapped Community Agreements) and JIA (the Jigsaw Interactive Agent) - each implementing different levels of alignment between LLM capabilities and collaborative learning frameworks. Both partners process spoken discourse from small-group interactions to provide support. CoBi uses a fine-tuned RoBERTa model that is trained to identify discourse behavior related to Community Agreements (e.g., moving thinking forward, being respectful, and committed to the community). Ongoing work is supporting the translation of this model to a generative LLM that can bring in custom classroom agreements and enables more contextual feedback to students. JIA is an LLM-based partner that generates real-time feedback based on multi-turn context windows, and is guided by principles of Productive Uncertainty to encourage elaboration, sense-making, and collaboration among students participating in small-group work. JIA was developed using three unique approaches: (1) discriminative models with rule-based thresholds and hard-coded feedback, (2) prompting an LLM with descriptions of our framework and desired types of feedback, and (3) a hybrid model that uses both rule-based outputs and framework descriptions to allow an LLM to be grounded in both pedagogical heuristics and generative LLM capabilities. All agents were evaluated using classroom data combining quantitative and qualitative performance metrics.

Results

Across agents, we observe clear trade-offs between interpretability, contextual depth, and pedagogical alignment depending on the dual use and roles of discriminative models and LLMs. Traditional fine-tuned models provide stable performance but limited adaptability to novel contexts. LLM-based agents achieve greater contextual sensitivity (identifying when students were reasoning well or exploring connecting ideas from a larger contextual window), but occasionally produce feedback that is not entirely pedagogically grounded. Early classroom trials indicate that framework-aligned prompting substantially improves feedback quality and relevance. For example, the hybrid JIA partner harnesses both rule-based knowledge and LLM-based contextual conversation information to intervene in a group that was evaluating gaming moderation systems by giving them support in problematizing, saying, "What makes you think the paid moderators approach is the most effective? Are there any potential downsides to this approach?". In response, students elaborated on the pros and cons of choosing AI + paid moderators as the best approach.

Significance

This work demonstrates both the promise and the necessity of theory-aligned AI for assessing and supporting collaborative learning. By embedding frameworks like Productive Uncertainty and CPS into model objectives, we move toward AI partners that not only recognize collaboration through defined constructs but that can also participate meaningfully in supporting it. Our findings suggest that combining LLMs' contextual strength with principled pedagogical alignment can enhance the assessment and support of collaborative discourse, offering methods for the next generation of AI partners in computer-supported collaborative learning environments.

Poster 4: Embedding AI partners within curricular routines

Jeffrey Bush Tayne, Greg Benedis Grab, Quentin Biddy, Thomas Breideband, Michael Alan Chang, William R. Penuel, Tamara Sumner

Objectives

This study investigates how curriculum routines can mediate the purposeful integration of AI-Partner tools into secondary classrooms in ways that support collaborative learning, equity, and student agency. Specifically, we focus on understanding the interactions between routines and AI tools in the context of co-designed AI partner, CoBi (Chang et al. 2024, Breideband et al., 2023), developed to help students establish and revisit community agreements that govern classroom collaboration (Bush, et al., 2025).

Perspectives

We draw on sociocultural theories that frame learning as a collective enterprise, where heterogeneity of ideas and participation is central to equity (Rosebery et al., 2010; Paris, 2012). At the same time, AI technologies are predisposed to convergent behaviors, reinforcing dominant patterns in their training data (Madaio et al., 2022). This creates tensions between equity goals and algorithmic bias. To address this, we developed curriculum routines, repeatable structures of teacher and student activity, with the intent to help students and teachers span boundaries in classrooms that help them engage with the AI in critical and agentic ways while avoiding the potential harms of AI's convergence bias. We analyze how these routines act as boundary objects (Akkerman & Bakker, 2011) that can mediate student transitions to productive reflection on their collaboration (meta-collaboration) and agentic critique of AI's accuracy.

Methods

The research draws on a multi-year partnership with teachers in a suburban/rural U.S. district. During the 2024-2025 school year, 11 teachers in various co-design and implementation focused PLCs implemented iSAT curriculum with CoBi. Using data from video of middle-grade (12-14 year old) classroom implementations, we employed thematic analysis (Braun & Clarke, 2006) to identify moments related to the research questions. We report on an instrumental case study (Stake, 1995) of Ms. Tessa's (pseudonym) classroom, one of several case studies in subsequent research.

Results

Teachers using CoBi predominantly enacted routines with high integrity, frequently adapting them in expansive ways aligned with goals. They often leveraged mistakes productively, incorporated student ideas, and reoriented routines toward new disciplinary aims. Though occasional disruptions occurred, overall enactments showed CoBi malleability for supporting collaboration and reflective learning. Despite this, opportunities for CoBi and the routines to support critical agency were rare; therefore, we present an instrumental case study highlighting an instance where CoBi did support this critical agency.

The case vignette illustrates how curriculum routines could function as a boundary object, mediating between the AI partner's algorithmic classifications and students' lived experiences of collaboration. In Ms. Tessa's classroom, the community agreement revisiting routine prompted students' initial reflections about how they collaborated and student critique about the AI system's judgements. By projecting CoBi's AI-generated sentence cloud of student talk and associated confidence intervals (for ASR and semantic classification), the teacher repositioned the AI's classifications as provisional, framing them as "proto-judgements" open to critique. This encouraged students to exercise critical agency, weighing the AI's labels against their own interpretations of respectful or productive talk. Rather than passively adopting system outputs, students engaged in meta-collaborative reflection, using the AI as a mirror to surface features of their teamwork that might otherwise remain invisible. The routine anchored dialogue while the tool generated new representations of collaboration that could be interrogated, sustaining an extended class conversation about norms, practices of equitable group work and students' lived experiences of collaboration.

Significance

This study shows how AI partners like CoBi can work with curriculum routines to present new representations of collaboration that anchor reflection and extend fleeting student observations into sustained dialogue. Embedded in curricular activity systems—professional learning, instructional materials, and classroom routines—CoBi supported high-integrity enactments, diverse adaptations, and productive use of mistakes and student ideas. Yet opportunities for critical agency were less frequent, requiring careful teacher framing. This research shows how CoBi's outputs can become provisional "proto-judgements," functioning as a boundary object that students could

 © ISLS

question, fostering meta-collaborative reflection. While critical agency was rare, the case study demonstrates the possibilities for designers to support teachers as they work to 1) balance AI's convergent tendencies with the need to value divergent contributions, and to 2) catalyze meta-collaborative reflections that sustain purposeful, collaborative learning.

Poster 5: Professional learning for teachers: Addressing triple innovation challenge

Collette Heskett, Nga Hoang, Mon-Lin Monica Ko, Julia Eden, William R. Penuel

Objectives

This poster presents a professional learning (PL) program designed to support the implementation of an AI-partner–embedded curriculum that promotes AI literacy. The program prepared teachers across disciplines to address a triple innovation challenge: developing their own AI literacy, becoming familiar with AI-enabled tools, and adopting student-centered pedagogical approaches.

Perspectives

Teachers face a challenge: both they and their students have unlimited access to AI tools, yet neither group has had opportunities to develop AI literacy (Touretzky et al., 2022). Instructing this new topic and embedding it into the curriculum, without agreed-upon standards, is challenging.

On iSAT, we designed curriculum-linked PL for secondary teachers across content areas to address a triple innovation challenge: teaching about AI, using AI-enabled tools, and adopting student-centered pedagogies. By leveraging teacher's diverse disciplinary perspectives, we explored both the technical and ethical facets of AI literacy while creating a testbed for adaptations of the same unit (Ko et al., 2023). Our PL design supported teachers in addressing the challenge of implementing the Moderation Unit—an AI-literacy unit that incorporates CoBi, an AI agent designed to facilitate classroom collaboration.

Research shows effective PL in STEM focuses on high-quality instructional materials (Lynch et al., 2019). Integrating curriculum and AI tools created both opportunities and challenges, especially when the curriculum itself teaches AI Literacy (Tan et al., 2025).

Methods

Seven secondary grade teachers from two neighboring U.S. school districts participated in a week-long (32-hour) PL workshop in Summer 2024. They represented diverse disciplines, including history, English Language Arts (ELA), math, and design thinking.

For this analysis, we drew on three PL activities: teachers experiencing the unit in a "student hat," interacting with CoBi in a "student hat," and planning for adaptations. The first activity supported AI literacy development; the second familiarized teachers with CoBi; the third allowed them to apply what they learned. Together, these emphasized student-centered learning and integration of AI literacy and tools.

Results

Teachers engaged with the triple innovation challenge by adapting instructional materials, planning to integrate CoBi, and fostering student-centered approaches.

First, teachers shared adaptations they planned to make to the Moderation unit to foster critical perspectives on AI within their disciplines. A math teacher, for example, planned to teach the unit as written while connecting the section on AI confidence scores to mathematical concepts. A middle school ELA teacher went further, designing an additional lesson on generative AI to help students analyze bias in the news and media. These adaptations showed how teachers linked AI literacy to their own disciplinary expertise and curricular goals.

Teachers also expressed enthusiasm for CoBi as a classroom tool. One noted it provided "100% positive reinforcement," while another emphasized that because CoBi is built on agreements set by students, "kids can expect what is going to happen and be able to relax in the environment." Such observations highlight CoBi's potential to encourage constructive participation and strengthen classroom community.

Finally, teachers demonstrated readiness to elevate student expertise in their implementation plans. For example, a high school ELA teacher described using driving question boards to give students more opportunities to generate their own questions, fostering authentic and student-centered conversations.

Significance

Our analysis revealed how the design of the PL workshop enabled teachers to develop their own AI literacy skills while incorporating them into their curricula. This provides insight into commitments needed for AI literacy curriculum and addresses the triple innovation challenge.

Poster 6: Designing and Refining AI Literacy Instructional Materials

William R. Penuel, Nga Hoang, Mon-Lin Monica Ko, Collette Heskett

Objectives

This poster describes the design and refinement of instructional materials supporting students' AI literacy. We share our co-design work with middle school youth and educators, and how implementation data has informed revisions to a unit on how computers learn from data. Our conception of AI literacy extends beyond model mechanics to include evaluating data, models, human input, and context shaping machine learning systems (TeachAI & CSTA, 2024).

Perspectives

We used a storyline model to center student learning on problems rooted in experience or community priorities (Penuel et al., 2024). Co-design teams—including youth—refined goals, selected problems, and built coherent lesson sequences. Iterative refinement addressed diverse teacher and student contexts, with varied sites reducing "hothouse" risks (Fishman & Krajcik, 2003) and professional learning supporting educator sensemaking.

Methods

Our project followed a Design-Based Implementation Research (DBIR) approach (Fishman et al., 2013). Goals and focal problems were negotiated with students, teachers, and district leaders. We adapted storyline co-design approaches (Penuel et al., 2022) to balance AI literacy goals with connections to students' lived experiences. Here, we provide a design narrative—a retrospective account of design choices informed by observations, student learning data, and teacher feedback (Gravemeijer & Cobb, 2013).

Results

Negotiating a focus for the unit

We modified the co-design process to incorporate youth voice and to exercise more agency in defining learning goals. Previously, youth contributed only through surveys, but here we recruited young people—primarily youth of color—from partner districts to help select the anchoring problem and shape the storyline. Survey data still informed the team, but classroom teachers, afterschool educators, and youth negotiated the unit's goals and design challenge.

Co-adaptation of the units

Because we recruited teachers across disciplines to enact the units, including STEM, science, mathematics, and English language arts, we recognized that the units could not be implemented as is, but would need to be adapted. Therefore, we adapted the professional learning model for storyline units (see McNeill et al., 2022) to include opportunities for teachers to review and identify how the unit's learning goals intersect with standards they typically address in their discipline. We also provided time for teachers to discuss and plan modifications to slides.

Iterative refinement of the units

Initial implementation in two 8th-grade classrooms (2023–2024) showed partial success. The unit was too long and often taught after state tests, limiting discussions of AI ethics. Student work showed awareness of algorithmic bias, but depth varied. We revised the unit by shortening it, focusing on AI and human moderation in gaming spaces, and strengthening lessons on bias in AI model training. These revisions improved usability and increased some teachers' confidence in implementing units.

Significance

Our project highlights challenges in developing instructional materials for AI literacy: co-designers need sufficient literacy themselves, units must adapt to multiple subject areas, integrating AI tools into instruction requires support, and assessments must guide iteration. While DBIR provides a framework for addressing these issues, our work shows that ingenuity and flexibility are essential.

 © ISLS

References

Akkerman, S. F., & Bakker, A. (2011). Boundary Crossing and Boundary Objects. *Review of Educational Research*, 81(2), 132-169. https://doi.org/10.3102/0034654311404435 (Original work published 2011)

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2), 77-101.

Breideband, T., Bush, J., Chandler, C., Chang, M., Dickler, R., Foltz, P., ... & D'Mello, S. (2023, October). The Community Builder (CoBi): Helping students to develop better small group collaborative learning skills. In Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing (pp. 376-380).

Bush, J. B., Biddy, Q., Chang, M. A., Grab, G. B., Breideband, T., Sumner, T., ... & D'Mello, S. (2025). Curriculum Routines and a Community Builder AI Agent: Synergy to Mediate the Convergent Tendencies of AI. In *Proceedings of the 18th International Conference on Computer-Supported Collaborative Learning-CSCL 2025*, pp. 614-616. International Society of the Learning Sciences.

Chang, M. A., Wong, R. Y., Breideband, T., Philip, T. M., McKoy, A., Cortez, A., & D'Mello, S. K. (2024). Co-design Partners as Transformative Learners: Imagining Ideal Technology for Schools by Centering Speculative Relationships. *In Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1–15. https://doi.org/10.1145/3613904.3642559

D'Mello, S. K., Biddy, Q., Breideband, T., Bush, J., Chang, M., Cortez, A., Flanigan, J., Foltz, P. W., Gorman, J. C., Hirshfield, L., Ko, M.-L. M., Krishnaswamy, N., Lieber, R., Martin, J., Palmer, M., Penuel, W. R., Philip, T., Puntambekar, S., Pustejovsky, J., … Whitehill, J. (2024). From learning optimization to learner flourishing: Reimagining AI in Education at the Institute for Student-AI Teaming (iSAT). *AI Magazine*, 45(1), 61–68. https://doi.org/10.1002/aaai.12158

Dowell, N. M. M., Nixon, T. M., & Graesser, A. C. (2019). Group communication analysis: A computational linguistics approach for detecting sociocognitive roles in multiparty interactions. *Behavior Research Methods*, 51(3), 1007–1041. https://doi.org/10.3758/s13428-018-1102-z

Fishman, B. J., & Krajcik, J. S. (2003). What does it mean to create sustainable science curriculum innovations? A commentary. *Science Education*, 87(4), 564–573.

Fishman, B. J., Penuel, W. R., Allen, A.-R., Cheng, B. H., & Sabelli, N. (2013). Design-Based Implementation Research: An emerging model for transforming the relationship of research and practice. In *B. J. Fishman, W. R. Penuel, A.-R. Allen, & B. H. Cheng (Eds). Design-based implementation research.* National Society for the Study of Education Yearbook, 112(2), 136–156. https://doi.org/10.1177/016146811311501415

Golumbic, Y. C., Adler, I., Gutman, N., Gueta, T., & Ben-Ami, F. (2025). Becoming a facilitator through co-design: A community of practice approach for facilitator engagement in a youth-centric citizen science initiative. *International Journal of Science Education, Part B*, 15(3), 493–508. https://doi.org/10.1080/21548455.2025.2488406

Graesser, A. C., Fiore, S. M., Greiff, S., Andrews-Todd, J., Foltz, P. W., & Hesse, F. W. (2018). Advancing the Science of Collaborative Problem Solving. *Psychological Science in the Public Interest*, 19(2), 59–92. https://doi.org/10.1177/1529100618808244

Gravemeijer, K., & Cobb, P. (2013). Design research from a learning design perspective. In *J. van den Akker, K. Gravemeijer, S. E. McKenney, & N. Nieveen (Eds.), Educational design research* (pp. 73–112). Routledge.

Hegedus S. & Kaput, J. (2003). *The Effect of a SimCalc Connected Classroom on Students' Algebraic Thinking. International Group for the Psychology of Mathematics Education*, Paper presented at the 27th International Group for the Psychology of Mathematics Education Conference Held Jointly with the 25th PME-NA Conference (Honolulu, HI, Jul 13-18, 2003), v3 p47-54

Ko, M.-L. M., Cortez, A., & Penuel, W. R. (2023). Curriculum as seed: Designing and supporting the use of open education resources to promote teacher agency and innovation. In *Blikstein, P., Van Aalst, J., Kizito, R., & Brennan, K. (Eds.), Proceedings of the 17th International Conference of the Learning Sciences* (pp. 1014-1017). International Society of the Learning Sciences.

Krumm, A. E., Penuel, W. R., Pazera, C., & Landel, C. (2020). Measuring equitable science instruction at scale. In *M. Gresalfi, I. S. Horn, N. Enyedy, H.-J. So, V. Hand, K. Jackson, S. E. McKenney, A. Leftstein, & T. M. Philip (Eds.), Proceedings of the International Conference of the Learning Sciences* (Vol. 4, pp. 2461–2468). International Society of the Learning Sciences.

Lynch, K., Hill, H., Gonzalez, K., & Pollard, C. (2019). Strengthening the research base that informs STEM instructional improvement efforts: A meta-analysis. *Educational Evaluation and Policy Analysis*, 41(3), 260–293. https://doi.org/10.3102/0162373719849044

Madaio, M., Egede, L., Subramonyam, H., Wortman Vaughan, J., & Wallach, H. (2022). Assessing the fairness of ai systems: Ai practitioners' processes, challenges, and needs for support. In *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW1), 1-26.

Manz, E., Suárez, E. (2018) Supporting Teachers to Negotiate Uncertainty for Science, Students, and Teaching. *Science Education* 102, no. 4 (2018): 771–95. https://doi.org/10.1002/sce.21343

Mawasi, A., Penuel, W. R., Cortez, A., & McKoy, A. (2023). "They were learning from us as we were learning from them": Perceived experiences in co-design process. *Mind, Culture, and Activity*, 30(3-4), 191–208. https://doi.org/10.1080/10749039.2023.2246039

McNeill, K. L., Affolter, R., & Reiser, B. J. (2022). Anchoring science professional learning in curriculum materials enactment: Illustrating theories in practice to support teachers' learning. In *A. C. Superfine, S. Goldman, & M.-L. M. Ko (Eds.), Teacher learning in changing contexts: Perspectives from the learning sciences* (pp. 47–68). Routledge. https://doi.org/10.4324/9781003097112-5

OECD (2025). Empowering learners for the age of AI: An AI literacy framework for primary and secondary education (Review draft). OECD. Paris. https://ailiteracyframework.org

OpenSciEd. http://www.openscied.org/. Accessed: 2025-02-12.

Penuel, W. R., Henson, K., Bracey, Z. B., VIck, N., & Rivet, A. E. (2024). Strategies for designing standards-aligned instructional materials that connect to students' interests and community priorities. *The Science Teacher*, 91(5), 62–70. https://doi.org/10.1080/00368555.2024.2390547

Penuel, W. R., Pazera, C., Buck Bracey, Z., Edelson, D. C., Henson, K., Reiser, B. J., Sumner, T., Vick, N., Ryan, J., & Watkins, D. A. (2025). Organizing design and development research for relevance and impact. *Educational Designer*, 5(19), 1–29.

Reich, J., & Dukes, J. (2024). Toward a new theory of arrival technologies: The case of ChatGPT and the future of education technology after adoption. Unpublished manuscript. https://osf.io/preprints/edarxiv/x6vn7_v1

Roschelle, J., Knudsen, J., & Hegedus, S. (2010). From new technological infrastructures to curricular activity systems: Advanced designs for teaching and learning. In *Designs for learning environments of the future* (pp. 233–262). Springer. http://link.springer.com/chapter/10.1007/978-0-387-88279-6_9

Rosebery, A. S., Ogonowski, M., DiSchino, M., & Warren, B. (2010). "The coat traps all your body heat": Heterogeneity as fundamental to learning. *The Journal of the Learning Sciences*, 19(3), 322-357.

Stake, R. (1995). Case study research. Cham: Springer.

Star, S. L. (2010). This is Not a Boundary Object: Reflections on the Origin of a Concept. *Science, Technology, & Human Values*, 35(5), 601-617. https://doi.org/10.1177/0162243910377624 (Original work published 2010)

Star, S. L., & Griesemer, J. R. (1989). Institutional Ecology, `Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social Studies of Science*, 19(3), 387-420. https://doi.org/10.1177/030631289019003001 (Original work published 1989)

Stilgoe, J., Owen, R., & Macnaghten, P. (2013). Developing a framework for responsible innovation. *Research Policy*, 42(9), 1568–1580. https://doi.org/10.1016/j.respol.2013.05.008

Sun, C., Shute, V. J., Stewart, A., Yonehiro, J., Duran, N., & D'Mello, S. (2020). Towards a generalized competency model of collaborative problem solving. *Computers & Education*, 143, 103672. https://doi.org/10.1016/j.compedu.2019.103672

Tan, X., Cheng, G., & Ling, M. H. (2025). Artificial intelligence in teaching and teacher professional development: A systematic review. Computers and Education: Artificial Intelligence, 8, 100355. https://doi.org/10.1016/j.caeai.2024.100355

TeachAI, & CSTA. (2024). Guidance on the future of computer science in the age of AI. Authors.

Tierney, G., & Stamatis, K. (2024). Constraints and boundaries for collective agency in co-design. In *R. Lindgren, T. Asino, E. A. Kyza, C. K. Looi, D. T. Keifert, & E. Suárez (Eds.), Proceedings of the 18th International Conference of the Learning Sciences* (pp. 1614–1617). International Society of the Learning Sciences.

Touretzky, D., Gardner-McCune, C., & Seehorn, D. (2022) Machine Learning and the Five Big Ideas in AI. *International Journal of Artificial Intelligence in Education*, 33, 233–266. https://doi.org/10.1007/s40593-022-00314-1

Vahey, P., Knudsen, J., Rafanan, K., & Lara-Meloy, T. (2013). Curricular Activity Systems Supporting the Use of Dynamic Representations to Foster Students' Deep Understanding of Mathematics. In C. Mouza & N. Lavigne (Eds.), *Emerging Technologies for the Classroom: A Learning Sciences Perspective* (pp. 15–30). Springer. https://doi.org/10.1007/978-1-4614-4696-5_2

International Society of the Learning Sciences

Revisiting Wise and Schwarz's Provocations for CSCL: Partnerships for Transformation and Educational Change

Breneil Malcolm (Co-Chair), Penn State University, bxm5750@psu.edu
Marcela Borge (Co-Chair & Discussant), Penn State University, mbs15@psu.edu
Alyssa Wise (Discussant), Vanderbilt University, alyssa.wise@vanderbilt.edu
Marcelo Worsley (Discussant), Northwestern University, marcelo.worsley@northwestern.edu
Sebastian Simon (Hybrid Ambassador), University of Copenhagen, sas@psy.ku.dk
Tanner Vea, University of Washington, tannerv@uw.edu
Natalie Rae, Penn State University, natalie.rae@psu.edu
Martin Greisel, University of Augsburg, martin.greisel@uni-a.de
Mia Čarapina, Zagreb University of Applied Sciences, mia.carapina@tvz.hr
Ban Mouid Shiwalia, University of Copenhagen, ban.shiwalia@psy.ku.dk

Abstract: In 2017, Wise and Schwarz proposed eight provocations for the field of CSCL following their interactions and analyses with scholars within the field. We broaden the call to the entire LS community to once again pause, reflect, and discuss a vision for the next 10 years of opportunities, challenges, and impacts we seek to make. This collective ambition is not simply to be responsive to technological changes but also to inform their design and use. Towards this end, we invited researchers from the ICLS and CSCL communities to propose new provocations for us to consider and discuss that focus both on the changing technological landscape and the need to ensure that diverse perspectives and viewpoints are brought into the forefront of technology-enhanced collaboration in meaningful ways.

Introduction

As our world grows ever more reliant on digital technologies to communicate, share knowledge, and think together, and as AI rapidly reshapes how we work and learn, the ICLS and CSCL communities are poised to make critical contributions in ensuring these tools foster the health and wellbeing of people at multiple levels of scale. We can draw on decades of collective research to inform design and technological integration to prioritize supporting individual learning and cognition, helping groups develop the capacity to engage in higher-order collective thinking and innovation, and strengthening communication processes and social infrastructures that stitch together the fabric of societal cohesion. Towards this aim, we revisit Wise and Schwarz's (2017) paper on "Visions of CSCL: Eight provocations for the future of the field." In this paper, the authors challenged the CSCL community to reflect on its beginnings and identify what they believed to be the greatest challenges the field would face in the coming years. As we approach the 10th anniversary of the provocations article and face a changing educational landscape, we broaden the call to the entire LS community to once again pause, reflect, and discuss a vision for the next 10 years of opportunities, challenges, and impacts we seek to make. The collective ambition is to not simply be responsive to technological change, but also capable of informing and shaping these developments through [our] rich theoretical landscape." (Kaliisa et al., 2025, p. 16).

The need to focus on technology-enhanced collaboration

Globally, we are seeing how social media technologies are amplifying inequality, polarization, politics of hate, and a splintering of society. People often use technologies like ChatGPT to learn about new things, which is both exciting but also potentially problematic. Learners are finding ways to use generative AI to spark creativity, opening possibilities for CSCL to help shape the design of playful, creative, and culturally relevant learning experiences that foster higher-order thinking and collaborative habits of mind (Kafai & Morales-Navarro, 2023). This, however, does not address the ongoing risk that generative AI may reinforce existing inequities and deepen the digital divide if left to be adopted in the wild.

There is concern that the rise of generative AI tools will amplify existing inequities given findings that generative AI use is higher among children in higher socio-economic classes and with children who attend private schools than public schools (Hashem et al., 2025). Recent research also points to the dangers of weaker engagement in and self-regulation of learning activities (Fan et al., 2024) and reduced cognitive activity, recall, and ownership of tasks (Kosmyna et al., 2025) when using generative AI in unscaffolded ways. Thus those learners with access to more experienced teachers, who are in schools that follow informed technology integration, with access to paid subscriptions that provide higher-quality tools, and opportunities to use technology for

 © ISLS

knowledge creation work, these are the learners that will reap the benefits of technology. In many ways, using technology to mitigate inequity is at the heart of foundational work in CSCL, but the focus was on using technology to give learners agency, not to address systemic inequities.

The need to build bridges across ICLS and CSCL

The dearth of research in CSCL that focuses on systemic inequity or the needs of diverse populations is problematic if we are to help inform the design and use of technology-enhanced collaboration for globally diverse populations (Gomez et al., 2021). Without full awareness of the ways that technology design and use can negatively impact differing populations, we run the risk of amplifying existing patterns of interaction that may harm many of our learners. More researchers within the broader Learning Sciences community focus on different forms of equity, culturally meaningful STEM, and learner-centered research (Fong et al., 2023; Rajala et al., 2022; Sheridan et al., 2022). However, much of this research is US-centric and less focused on technological implications (Sawyer, 2022). CSCL researchers tend to be more focused on the technological implications of research and more globally distributed (Baker & Reimann, 2025; Kurtz et al., 2025). Thus, by bridging these communities, we can identify new grand challenges and discuss whether and how educational transformation should be a goal of the International Society of the Learning Sciences.

Towards this end, we invited researchers from the ICLS and CSCL communities to propose new provocations for us to consider and discuss that focus both on the changing technological landscape and the need to ensure that diverse perspectives and viewpoints are brought into the forefront of technology-enhanced collaboration in meaningful ways.

Provocation 1: As a cultural practice that sustains the conditions for democracy, collaborative learning embodies a societal responsibility that must be prioritized

Sebastian Simon, Mia Čarapina, Ban M. Shiwalia

Divisive tendencies are on the rise worldwide. Global collaborations and democratic institutions are under strain, and technological infrastructures increasingly amplify fragmentation (International IDEA, 2025). History shows that democracies have weakened before, yet recovery has almost always come from below, through collective civic movements that depend on citizens' ability to deliberate, coordinate, and act together (Riedl et al., 2025). While such collaborations always have been challenging, today, such capacities are further challenged by technological progress: mass surveillance and social credit systems foster self-censorship and suppress social movements in their roots (Qiang, 2019). Simultaneously, algorithmic personalization in social media (Zuboff, 2019) and the rise of AI systems marketed as companions promise frictionless interaction while quietly eroding the messy but vital experience of human collaboration. Symptoms such as the loneliness epidemic (King, 2018), erosion of core collaborative values such as respect and trust in education, and within other social settings, foreshadow their larger impacts on societies(Levitsky & Ziblatt, 2024).

If democratic resilience depends on the collective capacity of societies to rebuild social trust, negotiate shared meaning, and listen to each other through empathy and dialogue, then building these capacities in future citizens becomes an educational imperative. We argue that collaborative learning is not merely a pedagogical method, but a cultural practice that sustains the very conditions for democracy in an age of technological individualisation and division. Everyone learns through and to navigate within complex social systems, and computer support has become ubiquitous in that regard. As such, CSCL and related fields share a responsibility towards current and future citizens to help them collaborate and understand collaboration better.

The question is how this is to happen when CSCL generally focuses on small-scale, isolated interventions with artificial activities designed by researchers (Wise & Schwarz, 2017). For even a systematic, shared, theory-driven understanding of collaborative learning requires, at the outset, small controlled studies in which variables can be isolated and collaborative learning processes examined rigorously. In this process, findings must be iteratively discussed with practitioners to situate theoretical mechanisms within the realities of classroom practice, rather than relying on idealized models. Through such investigations, researchers can identify the mechanisms that genuinely support learning in CSCL and differentiate between activities that facilitate collaboration to then support practitioners. Previously labeled ineffective interventions should, in this regard, not be seen as failures but as analytical insights that inform the development of supportive techniques, e.g., scripting (Vogel et al., 2017). It is crucial to understand that these techniques are not to dictate students' thinking processes, but to structure interactions so that meaningful collaboration can emerge, while simultaneously promoting co-knowledge acquisition and collaborative competencies.

 © ISLS

Once we agree on foundational principles (which we clearly still lack), research can start being translated into practice through meaningful partnerships with K-12 schools, universities, community-based organizations, and corporate partners, through collaborative efforts and methods, such as design-based research (DBR). In this way, the small-scale studies provide the basis for sustainable and scalable CSCL interventions that are both theoretically grounded and practically relevant. Feedback from practitioners can then inform further research on these collaborative practices, bound to change in the contexts they are practiced in. The artificiality of experimental, collaborative activities is another interesting point, since CSCL can have another, direct impact through the design of its activities. These activities implement state of the art CSCL theory and practice, such as positive interdependence or knowledge of the intricate social processes, and as such present an interesting opportunity for practitioners to explore such settings as a learning opportunity for themselves. One could argue that this mutual exchange is, in itself, collaborative learning, if conducted with practitioners as actors and not mere providers of study settings. Likewise, practitioners hold valuable practical knowledge that can inform meaningful CSCL interventions. But not only do we need a mutual collaborative learning opportunity between practitioners and researchers (in punctual, isolated studies). If isolated studies are an issue, one might ask why don't we apply what we know about collaborative learning to innovate how we learn as a community and practitioners to conduct large multilab experiments? DBR can guide multilab experimental design efforts through gradual refinement of initial experiments before sharedly executed on a large scale. We argue that, if we truly are to embody collaborative principles in our research, multi-lab interdisciplinary collaborations should be the norm, interventions both close to practice and transmitting collaborative skills should be the norm, and sharing collaborative knowledge and values with participants should be the norm. We challenge CSCL researchers to set and document an example of such a collective effort, create and refine procedures, activities & material to conduct experiments that simultaneously constitute learning experiences for practitioners across countries and laboratories, coordinate and engage as an asynchronous provider of knowledge and CSCL practice. Interestingly, going down this difficult road of collaborative research will open another, potentially just as interesting: the possibility to study collaboration between CSCL researchers who have awareness on what successful collaboration requires and means. Furthermore, by building community, CSCL can serve as an example for other communities and rebuild the spirit of collaborative learning on the many levels it is challenged across the world today. Learning is no isolated endeavour that happens in schools. It happens through interactions everywhere and all the time, and it is intrinsically collaborative. If we want society to prevail, CSCL must itself set the example and collaborate across every spectrum of it.

Provocation 2: We need to develop solidarity with communities to redefine what we mean by educational change

Natalie Rae

It is necessary to redefine what educational change might mean in the context of the intensifying socioecological crisis and the rise of generative AI (GAI) tools. Thinking alongside elaborations of educational change, such as Law and colleagues' (2021) dimensions of scalability, I advance a notion of educational change that takes up Philip and Sengupta's (2020) urgent call to see theories of learning within historical and contemporary processes of imperialism. This reorientation underscores the importance of a shift from tool-centric collaboration learning to collective, justice-oriented approaches that foreground collaboration as solidarity with social movements for justice and self-determination. In light of the ecological assessments of GAI's impacts and ISLS 2026 conference theme, *'Partnering with Purpose,'* I engage with literature in and beyond LS and CSLC to call for the sociopolitical process of developing solidarity with communities – as co-thinkers and partners in the struggle for collective socioecological thriving.

GAI's positioning as an emerging technology for learning

Since Wise and Schwarz's (2017) publication, there has been a tremendous increase in research on GAI in education (Ng & Ho, 2025). The hype and hyperbole around GAI's capabilities and utility for learning have been met with both enthusiasm for possible futures and caution (Borge et al., 2024): with scholarship examining evidence of ongoing social (e.g., Selwyn, 2022) and ecological (Markelius et al., 2024) harm. Increased financial speculation and investment in the development of physical infrastructure to enable GAI — expanding demands on fossil fuel energy use through data center development (Edwards et al., 2025) — traces the colonialism's patterns of expropriation and exploitation (Birhane, 2020). As communities within the U.S. (see Memphis, Tennessee, in Kolodny, 2025) and the majority world (see Chilean water activists and Uruguayan citizens in Hao, 2025) contest the development of data centers in their locales, social movements for self-determination are of immediate concern for the communities by which technology and capital interests have impinged (Faniyi, 2024).

 © ISLS

Self-determination here refers not only to a group's claims to autonomy and having consequential governance, but also to a collective, relational praxis rooted in the cosmologies of indigenous and Afro-descended peoples (e.g., Walsh, 2015). In the next section, I unpack an example of one of such movements, in Boxtown South Memphis, a case that makes clear the stakes of committing to and reimagining educational change for learning designers and scientists who consider learning within emerging technologies. Where scholars in the learning sciences and CSCL communities have articulated how theories and designs for learning are accountable to their contemporaneous political contexts (e.g., Philip & Sengupta, 2020, Learning Towards Justice, 2025), notions of educational change must reflect this context.

A case study in self-determination in the age of AI-data centers: Boxtown, South Memphis

This provocation began by wanting to think about how the conversation about AI use makes any conceptual contact with community efforts for self-determination, rooted in the intellectual genealogies of politicized thought and action, and grew out of my inquiry into a historically Black town in South Memphis, Tennessee, Boxtown. In 2024, the former Electrolux factory in Memphis was transformed into a data center for Elon Musk's company xAI's product, Grok. Since the factory was taken over, community members have expressed concerns about the use of natural gas generators – more than were permitted – as exacerbating air-quality problems (Harris et al., 2025).

The Boxtown community in South Memphis represents a robust case of resistance and ecological justice in the face of industrial encroachment and environmental racism. The Boxtown community was formed by formerly enslaved Black people who built homes from abandoned boxcars, creating an independent community (Moore, 2021). Enduring decades of industrial pollution, including exposure to refinery emissions, coal ash, and steam plant operations, studies indicate significantly elevated cancer rates and other health disparities among residents. The Boxtown community, people, and ecology have been treated as expendable —an ostensible sacrifice zone (Juskus, 2023) — in the face of industrialism and expanding profit. Boxtown residents have been collectively pushing back against injustice, most recently their successful opposition to the Byhalia Connection Pipeline (First-Arai, 2022). My historical coverage is meant to underscore how these industrial intrusions are not incidental but part of the economic, political, and ideological work that renders lives as expendable to the will of industrial progress. And yet, these communities continue to resist, such as in the Memphis 1968 Sanitation Strike, considered a precursor to the contemporary environmental justice movement (League of Conservation Voters, 2022). The tools, strategies, and intellectual legacy of activism for self-determination is a collective, relational praxis rooted in the cosmologies of indigenous and Afro-descended peoples (e.g., Walsh, 2015). In each of these movements, the strategies of organizing people, co-developed priorities and tactics, the ways of partnering and coming together are forms of learning and action as much as the products of their efforts - an articulation for reparative justice—calling for ecological redevelopment, solar energy investment, and the restoration of Black and Indigenous relationships to land and place for themselves and communities like them. Secondly, it would be disingenuous to frame Boxtown's activism and calls for self-determination only as relevant within the contemporaneous concerns specific to xAI's Colossus 1 Data center operations, but rather part of a longstanding struggle of racially impacted industrialization.

How can we think about domain learning or collaboration when the learners themselves cannot breathe? My question is not a compositional flourish: I ask this with impatient sincerity, as I am but one of the growing number of Black parents who have children with respiratory conditions that can keep them out of school and threaten their wellness. What is the point of Computer Supported Collaborative Learning when the computational infrastructure is at the expense of Black peoples' lives and our ecologies' ability to thrive? For my thinking, Táíwò's (2022) concept of constructive politics: the creation of new social structures and relationships that redistribute power and resources. For Boxtown's demands go beyond inclusion in existing governance with regard to one technology—they seek radical transformation of political and economic systems. For CSCL, this case underscores the need to recognize community organizing as a form of learning, in which scientific inquiry, environmental monitoring, and advocacy are collaborative, pedagogical sites of transformative agency (Bang & Vossoughi, 2016). Incorporating such cases into CSCL discourse challenges the field to expand its unit of analysis from individual learners to communities and ecologies. It also calls for a redefinition of educational change—one that centers solidarity, reparations, and ecological justice as urgent and transformative learning outcomes.

Design implications for CSCL: Constructive politics and solidarity-centered learning

Addressing this context requires LS and CSCL scholars to treat the hyper-local scale of solidarity-centered learning as the scale for educational change. Drawing on Law et al.'s (2021) five dimensions of scalability—depth, sustainability, spread, ownership, and evolution—I call on CSCL to reimagine its design principles to support

justice-oriented learning, specifically in the context of social movements for self-determination in the global proliferation contexts of data centers. This involves moving beyond tool-centric models to embrace community-centered learning architectures, thereby emphasizing the crucial role of community involvement in educational design.

Constructive politics offers a framework for these transformations. It invites CSCL researchers and designers to build 'the rooms'—physical and conceptual spaces—where communities can imagine and enact just futures. This includes developing infrastructures that support learning as resistance, such as platforms for environmental monitoring, storytelling, and policy advocacy. Moreover, CSCL must adopt design principles that center on solidarity. This means engaging with peripheral contexts (Matuk et al., 2021), where affected communities are treated as co-designers for collective action in which we are personally accountable to. Ultimately, CSCL's role in educational change lies not in scaling tools, but in scaling justice. By aligning with social movements and ecological struggles, the field can contribute to learning environments that are not only collaborative but also liberatory.

Provocation 3: We need to bridge epistemological differences between LS and CSCL to cultivate more productive collaborations

Martin Greisel

Any call to collaborate between the Learning Sciences (LS) and Computer-Supported Collaborative Learning (CSCL) will be ineffective if it remains unclear how to bridge epistemological differences, which are rooted in different ontological stances and axiological values and norms (AOE). This claim rests on three assumptions: (a) Researchers differ in their ontologies and axiologies which, in turn, constrain their epistemic aims, the ideals or standards they value, and the processes they consider as producing reliable knowledge (see AIR framework; Barzilai & Chinn, 2024); (b) genuine collaboration requires some understanding of these ontological, epistemological, and axiological commitments; and (c) while LS is often associated with interpretative, design-based, or social justice-oriented traditions, CSCL spans a broader mix of approaches with a noticeable tendency toward analytical or computational methods.

The challenge, then, is not simply to connect two communities but to navigate differences that cut across both. Scholars from different communities may understand each other easily if they share similar commitments, whereas scholars within the same community might struggle if one seeks ideographic insights through interpretative reasoning grounded in situated meaning-making and the other seeks nomothetic explanation through an analytical, variable-based approach. The key issue is therefore less LS versus CSCL than dialogue across interpretative, analytical, design-based, and social justice-oriented epistemologies and their accompanying ontologies and axiologies.

Learning scientists already collaborate, but collaboration does not guarantee mutual understanding. They work within a shared organizational infrastructure of societies, conferences, committees, and journals (i.e., an architecture of collaboration; Shrum et al., 2007). Within it, researchers with different AOE stances work together through shared interfaces such as co-authored papers, creating trading zones in Galison's (1997) sense. Such collaborations often begin around common terms like *learning* or *collaboration*, yet these terms are frequently polysemous (Ceccarelli, 2001) and can conceal important differences in meaning.

Moving beyond the pidgin of the trading zone therefore requires a more explicit *shared vocabulary*. Understanding across ontological, epistemological, and axiological differences requires more than goodwill; it requires tools that support meta-level reflection on how we produce knowledge. A shared vocabulary would be a first step. This does not imply a single agreed definition of "learning," "collaboration," or "theory," but language that makes divergent definitions explicit and mutually intelligible. Without such linguistic infrastructure, researchers may talk past each other, mistaking epistemic differences for disagreements in substance, or ignore work they cannot relate to their own.

As a first step, to provide such a linguistic infrastructure and to make these differences visible, the community could develop a *taxonomy*. Such a taxonomy would classify researchers' ontologies (e.g., what is the nature of "learning"), the epistemic aims (e.g., understand meaning-making, build descriptive or predictive models, develop design principles, transform power relations), the standards of evidence (e.g., thick description & theoretical resonance, statistical robustness & generalizability, practical consequentiality & iterative justification, justice in the research process & liberatory assessment), and the knowledge-generating processes (e.g., grounded interpretation & co-interpretation with participants, experimental manipulation & reliable measurement, iterative design & pragmatic evaluation, reflexive documentation & community feedback; Barzilai & Chinn, 2024). In this sense, I extend Rummel's (2018) proposal for a taxonomy of ways to support collaboration with technology. Such an (extended) taxonomy would not prescribe what "counts" as valuable and credible

research; rather, it would act as a reflective scaffold, helping researchers situate their work transparently and recognize others' positions without reducing them to labels like "LS" or "CSCL."

One could imagine a lightweight tagging system for publications or conference papers, through which authors indicate their ontological, epistemological and axiological profile, their methods, and the phenomenon they study using shared categories from the taxonomy. For example, contributions could self-identify as interpretative with the aim of a deep understanding of meaning-making mechanisms in a specific context, or analytical with the aim of generalizable knowledge, design-based with the aim of developing artefacts that are useful in educational practice, or social justice-oriented with the aim of transforming existing power relations. Making these commitments explicit could help readers interpret arguments more generously and accurately and see more easily how different approaches complement one another in addressing complex learning phenomena.

By shifting the conversation from community boundaries to ontological, epistemological, and axiological transparency, we might cultivate more productive collaborations. Rather than asking how LS and CSCL can work together, we might ask how scholars with different commitments can build reciprocal understanding without forcing assimilation into a single dominant discourse. A community-wide taxonomy and shared vocabulary would be a pragmatic step toward that goal, encouraging intentional pluralism while reducing inadvertent fragmentation.

Provocation 4: None of this matters if we don't stop authoritarianism

Tanner Vea

There was a time when I used to go on the internet to catch up with friends, post pictures of my mundane life, and yes—learn collaboratively. Compared to a couple of years ago, when I stood in front of my colleagues in the learning sciences to warn them of fascism (Vea, 2023), we in the United States have lost further ground. It is becoming increasingly common in the US to see videos of masked secret police making children stand naked in the cold street at night and whisking people away in unmarked vans (Hernandez, 2025). There are revelations about a group chat of young conservatives, the future of our ruling party, joking about gas chambers (Beeferman & Ngo, 2025). Rising authoritarianism is not unique to the United States (Gorokhovskaia & Grothe, 2025). The protection of tenure and the idea of academic freedom used to be pillars of the academy as well, but these too are weakening as the US government makes demands for how educational institutions are run, who is hired, and what is taught (Helsel, 2025). Fascism seeks to control education to limit dissent over the installation of extreme social hierarchy (Stanley, 2018). This reality is reinforced on the internet as we see post after post about professors put on leave or fired for speaking against hate and authoritarianism *in the wrong ways*. Every once in a while, I wonder about whether I have done so *in the right ways* and might therefore be safe. Then I remember that this, too, is part of the long slide into authoritarianism. The fear of unthinkable things, now made thinkable, shapes every other thought.

Arcing through the rising tide of authoritarianism is the erosion of truth (Barzilai & Chinn, 2020). We cannot understand our current epistemic challenges without understanding power (Finchelstein, 2022). Traditional gatekeeping structures have weakened, and now "populist expertise" (Marwick & Partin, 2024) in digital social (and learning) spaces fills the void with conspiracy theories. Leaders, experts, and lay people mingle online in ways that once seemed to hold promise for a new democratization (Barlow, 1996) but now has become a vector of disinformation (Starbird et al., 2023). If social media is an important site of learning (provocation 7; Wise and Schwarz, 2017), is this not the most prominent use case of computer-supported collaborative learning in our current age?

As these events play out and impact us on a global scale, we must ask if the learning sciences can inform these debates or use the scholarly enterprise to help shape alternative possibilities. There are positive cases where design and analysis illuminate, in my view, the kinds of knowledge we now need (e.g., Jurow et al., 2016; Russo & Blikstein, 2023). But I also think of the senior scholar who told me that activism is not science or another who disrupted a student's presentation for calling out the implication of technology in mass murder. Those voices seem to say: How dare we have anything to say about the world we live in? There will be consequences for being a learning scientist who steps out of line. In the meanwhile, more laws are broken. More bodies pile up. More pain, and less humanity, on my screen.

I do believe that the CSCL community and the broader learning sciences community must work together now for a field that is worthy of our moment. We cannot do so by policing each other into asking smaller and smaller questions about learning, as if retreating from the conditions of our lives will protect the sanctity of the scientific endeavor. I became a learning scientist to build on the work of others who illuminated that learning is always political (e.g., Esmonde & Booker, 2017; Gutiérrez & Rogoff, 2003; Nasir & Hand, 2006; Philip et al., 2017; The Politics of Learning Writing Collective, 2017). The "what" and "how" of learning cannot be

meaningfully separated from questions about what is good and right and who has access to it (Philip et al., 2018). Power needs to be brought into our understanding of every community of practice (Curnow, 2022).

For some time, these considerations have marked a kind of epistemic divide in our fields—between those who seek to understand how power shapes our learning and those who hope for a vision of science that seems cleaner or less polluted by our commitments to the world and each other (Learning Towards Justice Team, 2025). If ever there was a moment to heal that divide it is now. If, as Cole (1994) offered, context is "that which weaves together" (p. 93), then the context of rising authoritarianism is weaving together everything we are doing in the learning sciences. It is literally shaping who is in the room, versus who is online, versus those who are absent altogether. It is shaping the questions we are allowing each other to ask and answer. It is the fear of the previously unthinkable that now enters my every thought. If we do not use this moment to stand up for one another, we may lose the ability to conduct rigorous learning research that seeks equitable learning possibilities for *every* child. Let us use this historical moment to stretch beyond our narrow questions to improve the knowledge-building conditions we live and work within, by directing rigorous inquiry toward answering how we got into this mess, and how we will get out. Our fields' design orientation provides unique tools for thinking pragmatically, not only theoretically, about how our current arrangements of power and knowledge can be shifted. We should build, investigate, and protect spaces of possibility for refusing extreme social hierarchy, exclusion, and censorship. We should also use our professional associations and scholarly ways of working to speak back to authoritarianism with whatever remaining public credibility we have. Harder still, we need to ask ourselves how expectations to seek government funding from authoritarian regimes may implicate us in the very kinds of social engineering we hope to resist. Only in solidarity will we succeed.

Engaging the community

Each of our four provocations presents pathways for our LS and CSCL fields to pause, reflect, deconstruct, and reconstruct pathways forward. Provocation 1 invites us to embody and prioritize the societal responsibility that sustains our conditions for democracy; if democracy, then, is hinged on its cultural practice as collaborative learning, multi- and interdisciplinary onto-epistemologies are inherent to this collective effort of LS and CSCL, Simon, Čarapina, and Shiwalia assert. Provocation 2 incites an epistemological shift in how we construe educational change as a mechanism of facilitating solidarity, accounting for the sociopolitical contentions of GAI. In the argument for constructive politics, Rae makes the call for a reimagining of design principles that may support justice-oriented learning. Provocation 3 beckons us to embrace epistemological differences between LS and CSCL as a precursor to more productive collaborations. Greisel contends for a more explicit shared vocabulary to ground this linguistic infrastructure. Provocation 4 invokes the exigence of stopping authoritarianism, a call for solidarity. Vea advocates for the explicit rejection of epistemological policing and monistic definitions of knowledge construction and inquiry to mend the divides between LS and CSCL. Interestingly, together these provocations make a case for a plan forward. The synthesis of these papers push our collective fields to begin the work of translating our research for the public in ways that prioritize designing for justice-oriented learning. In doing so, we can draw on our epistemological differences as a source of strength if we can develop a research community with epistemological humility that recognizes the limits of our own methods and values alternative methods as different lenses. Each methodological lens has its own depth, scale, and ability to see and uncover diverse phenomena. If we can find ways to bridge our epistemological differences, we can work to create pedagogies grounded in dialogue, where engaging difference becomes a necessary precursor not only to stronger collaboration, but to more critical, humanizing, and socially transformative learning. As our final provocation warns, solidarity is a necessary condition for resisting authoritarianism. As learning scientists, we must protect the ability of diverse scholars to think, investigate, speak, dissent, and engage in collective decision-making without coercion or control, while also recognizing that education is never politically neutral. In this sense, the provocation echoes Baldwin's (1998) warning that "people who shut their eyes to reality simply invite their own destruction." It makes clear that we are responsible not only for supporting those who name and challenge existing power structures, but also for doing that work ourselves.

These provocations are offered not as conclusions but as openings for conversation, critique, and co-creation within and across the CSCL and LS communities. We invite you to respond, challenge, extend, and reimagine them in light of your own work and perspectives, and to propose new provocations that surface emerging questions, contexts, and commitments not represented here. Our aim is to foster new lines of inquiry and sustain a shared dialogue on how LS and CSCL research can shape the technologies and practices of collaboration in the decade ahead, with attention to both our shared foundations and the diverse values, theories, and commitments that shape our collective practice.

References

Baker, M., Reimann, P. (2025). Editorial notes: Core issues in CSCL research. *International Journal of Computer-Supported Collaborative Learning* 20, 165–170 https://doi.org/10.1007/s11412-025-09449-4

Baldwin, J. (1998). *Collected essays: Notes of a native son / Nobody knows my name / The fire next time / No name in the street / The devil finds work / Other essays* (T. Morrison, Ed.). Library of America.

Bang, M., & Vossoughi, S. (2016). Participatory Design Research and Educational Justice: Studying Learning and Relations Within Social Change Making. *Cognition and Instruction*, *34*(3), 173–193. https://doi.org/10.1080/07370008.2016.1181879

Barlow, J. P. (1996). *A declaration of the independence of cyberspace*. Electronic Frontier Foundation. https://www.eff.org/cyberspace-independence

Barzilai, S., & Chinn, C. A. (2020). A review of educational responses to the "post-truth" condition: Four lenses on "post-truth" problems. *Educational Psychologist*, *55*(3), 107–119. https://doi.org/10.1080/00461520.2020.1786388

Barzilai, S., & Chinn, C. A. (2024). The AIR and Apt-AIR frameworks of epistemic performance and growth: Reflections on educational theory development. *Educational Psychology Review*, *36*(3), 91. https://doi.org/10.1007/s10648-024-09927-5

Borge, M., Smith, B., & Aldemir, T. (2024). Using generative AI as a simulation to support higher-order thinking. *The International Journal of Computer Supported Collaborative Learning*, *19*(4), 479-532. https://doi.org/10.1007/s11412-024-09437-0

Beeferman, J., & Ngo, E. (2025, October 14). 'I love Hitler': Leaked messages expose Young Republicans' racist chat. *Politico*. https://www.politico.com/news/2025/10/14/private-chat-among-young-gop-club-members-00592146

Birhane, A. (2020). Algorithmic colonization of Africa. *SCRIPT-Ed*, *17*(2), 389–409. https://doi.org/10.2966/scrip.170220.389

Ceccarelli, L. (2001). *Shaping science with rhetoric: The cases of Dobzhansky, Schrodinger, and Wilson*. University of Chicago Press.

Cole, M. (1994). A conception of culture for a communication theory of mind. In D. R. Vocate (Ed.), *Intrapersonal communication: Different voices, different minds* (pp. 77–98). Erlbaum.

Curnow, J. (2022). Resituating situated learning within racialized and colonial social relations. *Mind, Culture, and Activity*, *29*(4), 301–315. https://doi.org/10.1080/10749039.2022.2066133

Edwards, D., Cooper, Z. G. T., & Hogan, M. (2025). The making of critical data center studies. *Convergence: The International Journal of Research into New Media Technologies*, *31*(2), 429–446. https://doi.org/10.1177/13548565231224157

Esmonde, I., & Booker, A. (Eds.). (2017). *Power and privilege in the learning sciences: Critical and sociocultural theories of learning*. Routledge.

Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., ... & Gašević, D. (2025). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology*, *56*(2), 489–530. https://doi.org/10.1111/bjet.13544

Faniyi, O. (2024, February 27). An African feminist manifesto. *The Republic*. https://republic.com.ng/february-march-2024/an-african-feminist-manifesto/

Finchelstein, F. (2022). *A brief history of fascist lies*. University of California Press.

First-Arai, L. (2022). *Lesson 7: Do it again* (No. 7) [Broadcast]. https://www.selc.org/broken-ground-season-5/

Fong, M. M., DeLiema, D., Flood, V. J., & Walker-van Aalst, O. (2023). Contesting sociocomputational norms: Computer programming instructors and students' stancetaking around refactoring. *International Journal of Computer-Supported Collaborative Learning*, *20*, 79–119. https://doi.org/10.1007/s11412-023-09392-2

Galison, P. (1997). *Image and logic*. University of Chicago Press.

Gomez, K., Gomez, L. M., & Worsley, M. (2021). Interrogating the role of CSCL in diversity, equity, and inclusion. In U. Cress, C. Rosé, A. F. Wise, & J. Oshiam (Eds.), *International handbook of computer-supported collaborative learning* (pp. 103–119). Cham: Springer International Publishing.

Gorokhovskaia, Y., & Grothe, C. (2025). *Freedom in the world 2025: The uphill battle to safeguard rights*. Freedom House. https://freedomhouse.org/sites/default/files/2025-03/FITW_World2025digitalN.pdf

Gutiérrez, K. D., & Rogoff, B. (2003). Cultural ways of learning: Individual traits or repertoires of practice. *Educational Researcher*, *32*(5), 19–25. https://doi.org/10.3102/0013189X032005019

Hao, K. (2025). *Empire of AI: Dreams and nightmares in Sam Altman's OpenAI*. Penguin Press.

Harris, B., Gerberg, J., & Gosk, S. (2025). Up against Musk's Colossus supercomputer, a Memphis neighborhood fights for clean air. *NBC News*. https://www.nbcnews.com/news/us-news/musk-xai-colossus-supercomputer-boxtown-memphis-tennessee-rcna206242

Hashem, Y., Esnaashari, S., Onslow, K., Chakraborty, S., Poletaev, A., Francis, J., Bright, J. (2025). *Understanding the impacts of generative AI use on children: WP1 surveys*. The Alan Turing Institute. https://www.rivista.ai/wp-content/uploads/2025/06/1749085651577.pdf

Helsel, P. (2025, October 16). How universities are responding to the Trump administration's 'compact' for funding. *NBC News. https://www.nbcnews.com/politics/politics-news/universities-are-responding-trump-administrations-compact-funding-rcna238095*

Hernandez, C. (2025, October 1). Massive immigration raid on Chicago apartment building leaves residents reeling: "I feel defeated." *Chicago Sun-Times*. https://chicago.suntimes.com/immigration/2025/10/01/massive-immigration-raid-on-chicago-apartment-building-leaves-residents-reeling-i-feel-defeated

International IDEA. (2025). The Global State of Democracy 2025: Democracy on the Move (The Global State of Democracy, p. 332). International IDEA. https://www.idea.int/publications/catalogue/global-state-of-democracy-2025-democracy-on-the-move

Jurow, A. S., Teeters, L., Shea, M., & Van Steenis, E. (2016). Extending the consequentiality of "invisible work" in the food justice movement. *Cognition and Instruction*, *34*(3), 210–221. https://doi.org/10.1080/07370008.2016.1172833

Juskus, R. (2023). Sacrifice Zones. *Environmental Humanities*, *15*(1), 3–24. https://doi.org/10.1215/22011919-10216129

Kafai, Y. B. & Morales-Navarro, L. (2023). Twenty Constructionist Things to Do with Artificial Intelligence and Machine Learning. In Proceedings of Fablearn/Constructionism 2023. https://arxiv.org/abs/2402.06775

Kaliisa, R., Lopez-Pernas, S., Misiejuk, K., Damşa, C., Sobocinski, M., Järvelä, S., & Saqr, M. (2025). A topical review of research in computer-supported collaborative learning: Questions and possibilities. *Computers & Education*, *228*, 105246. https://doi.org/10.1016/j.compedu.2025.105246

King, M. (2018). Working to address the loneliness epidemic: Perspective-taking, presence, and self-disclosure. *American Journal of Health Promotion*, *32*(5), 1315–1317. https://doi.org/10.1177/0890117118776735c

Kolodny, L. (2025, July 17). *Elon Musk's xAI permits challenged by NAACP, environmental groups in Memphis*. CNBC. https://www.cnbc.com/2025/07/16/musks-xai-permits-challenged-by-naacp-environmental-groups-memphis.html

Kosmyna, N., Hauptmann, E., Yuan, Y. T., Situ, J., Liao, X. H., Beresnitzky, A. V., ... & Maes, P. (2025). Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for essay writing task. *arXiv preprint arXiv:2506.08872*, *4*.

Kurtz, M., Benabbou, A., Pons, C., & Broisin, J. (2025). Collaboration in virtual and remote laboratories for education: A systematic literature review. *International Journal of Computer-Supported Collaborative Learning*, 1-55. https://doi.org/10.1007/s11412-025-09454-7

Law, N., Zhang, J., & Peppler, K. (2021). Sustainability and scalability of CSCL innovations. In U. Cress, C. Rosé, A. F. Wise, & J. Oshima (Eds.), *International Handbook of Computer-Supported Collaborative Learning* (pp. 121–141). Springer International Publishing. https://doi.org/10.1007/978-3-030-65291-3_7

League of Conservation Voters. (2022, February 14). Environmental Justice: A Movement Led and Progressed by Black Trailblazers - Part 1. *League of Conservation Voters*. https://www.lcv.org/blog/environmental-justice-a-movement-led-and-progressed-by-black-trailblazers/

Learning Towards Justice Team. (2025). Avoiding equity detours on the way to a more justice-oriented learning sciences. *Journal of the Learning Sciences*, *34*(3), 368–402. https://doi.org/10.1080/10508406.2025.2468213

Levitsky, S., & Ziblatt, D. (2024). How democracies die. In T. Ball, R. Dagger, D. I. O'Neill, & J. Kirkpatrick (Eds.), *Ideals and ideologies* (12th ed.). Routledge.

Marwick, A. E., & Partin, W. C. (2024). Constructing alternative facts: Populist expertise and the QAnon conspiracy. *New Media & Society*, *26*(5), 2535–2555. https://doi.org/10.1177/14614448221090201

Matuk, C., DesPortes, K., & Hoadley, C. (2021). Conceptualizing context in CSCL: Cognitive and sociocultural perspectives. In U. Cress, C. Rosé, A. F. Wise, & J. Oshima (Eds.), *International Handbook of Computer-Supported Collaborative Learning* (pp. 85–101). Springer International Publishing. https://doi.org/10.1007/978-3-030-65291-3_5

Moore, R. E. (2021). *Boxtown: A forgotten piece of Memphis history*. Memphis Mirror. https://web.archive.org/web/20210624201115/https://memphismirror.com/boxtown-a-forgotten-piece-of-memphis-history/

Nasir, N. S., & Hand, V. M. (2006). Exploring sociocultural perspectives on race, culture, and learning. *Review of Educational Research*, *76*(4), 449–475. https://doi.org/10.3102/00346543076004449

Ng, S.-L., & Ho, C.-C. (2025). Generative AI in education: Mapping the research landscape through bibliometric analysis. *Information*, *16*(8), 657. https://doi.org/10.3390/info16080657

Philip, T. M., Bang, M., & Jackson, K. (2018). Articulating the "how," the "for what," the "for whom," and the "with whom" in concert: A call to broaden the benchmarks of our scholarship. *Cognition and Instruction*, *36*(2), 83–88. https://doi.org/10.1080/07370008.2018.1413530

Philip, T. M., Gupta, A., Elby, A., & Turpen, C. (2017). Why ideology matters for learning: A case of ideological convergence in an engineering ethics classroom discussion on drone warfare. *Journal of the Learning Sciences*, *27*(2), 183–223. https://doi.org/10.1080/10508406.2017.1381964

Philip, T. M., & Sengupta, P. (2021). Theories of learning as theories of society: A contrapuntal approach to expanding disciplinary authenticity in computing. *Journal of the Learning Sciences*, *30*(2), 330–349. https://doi.org/10.1080/10508406.2020.1828089

Qiang, X. (2019). The road to digital unfreedom: President Xi's surveillance state. *Journal of Democracy*, *30*(1), 53–67. https://doi.org/10.1353/jod.2019.0004

Rajala, A., Cole, M., & Esteban-Guitart, M. (2022). Utopian methodology: Researching educational interventions to promote equity over multiple timescales. Journal of the Learning Sciences, 32(1), 110–136. https://doi.org/10.1080/10508406.2022.2144736

Riedl, R. B., Friesen, P., McCoy, J., & Roberts, K. (2025). Democratic backsliding, resilience, and resistance. *World Politics*, *77*(1, suppl.), 151–177. https://doi.org/10.1353/wp.2025.a954440

Rummel, N. (2018). One framework to rule them all? Carrying forward the conversation started by Wise and Schwarz. *International Journal of Computer-Supported Collaborative Learning*, *13*(1), 123–129. https://doi.org/10.1007/s11412-018-9273-2

Russo, R., & Blikstein, P. (2023). Just asking questions: Can a far-right president turn agentic knowledge construction into political manipulation? *Information and Learning Sciences*, *124*(7/8), 197–220. https://doi.org/10.1108/ILS-10-2022-0118

Sawyer, R. K. (2022). An introduction to the learning sciences. In K. R. Sawyer (Ed.), The Cambridge handbook of the learning sciences, 3, (pp. 1–24).

Selwyn, N. (2022). The future of AI and education: Some cautionary notes. *European Journal of Education*, *57*(4), 620–631. https://doi.org/10.1111/ejed.12532

Sheridan, K. M., Zhang, X., & Konopasky, A. W. (2022). Strategic shifts: How studio teachers use direction and support to build learner agency in the figured world of visual art. Journal of the Learning Sciences, 31(1), 14–42. https://doi.org/10.1080/10508406.2021.1999817

Shrum, W., Genuth, J., & Chompalov, I. (2007). *Structures of scientific collaboration*. MIT Press.

Stanley, J. (2018). *How fascism works: The politics of us and them*. Random House.

Starbird, K., DiResta, R., & DeButts, M. (2023). Influence and improvisation: Participatory disinformation during the 2020 US election. *Social Media + Society*, *9*(2), 20563051231177943. https://doi.org/10.1177/20563051231177943

Táíwò, O. O. (2022). *Elite capture: How the powerful took over identity politics (and everything else)*. Pluto Press.

The Politics of Learning Writing Collective. (2017). The learning sciences in a new era of U.S. nationalism. *Cognition and Instruction*, *35*(2), 91–102. https://doi.org/10.1080/07370008.2017.1282486

Vea, T. (2023, June 15). *Presidential session*. International Society of the Learning Sciences Annual Meeting, Montreal, Canada.

Vogel, F., Wecker, C., Kollar, I., & Fischer, F. (2017). Socio-cognitive scaffolding with computer-supported collaboration scripts: A meta-analysis. *Educational Psychology Review*, *29*(3), 477–511. https://doi.org/10.1007/s10648-016-9361-7

Walsh, C. E. (2015). Decolonial pedagogies walking and asking. Notes to Paulo Freire from AbyaYala. *International Journal of Lifelong Education*, *34*(1), 9–21. https://doi.org/10.1080/02601370.2014.991522

Wise, A. F., & Schwarz, B. B. (2017). Visions of CSCL: Eight provocations for the future of the field. *International Journal of Computer-Supported Collaborative Learning*, *12*(4), 423–467. https://doi.org/10.1007/s11412-017-9267-5

Zuboff, S. (2019). Surveillance capitalism and the challenge of collective action. *New Labor Forum*, *28*(1), 10–29. https://doi.org/10.1177/1095796018819461

Hybrid facilitation plan

To facilitate accessibility and attendance for those who will not be able to attend the conference in person, especially accounting for the challenging political climate, we will offer a hybrid symposium. Towards that end, we will provide a variety of asynchronous, synchronous, online, and in-person opportunities for engagement before, during, and after the sessions.

Before the session

We will create a form of asynchronous engagement with the provocations through a website that includes a 30-second pitch of each provocation as a short video, along with its respective author(s) and written section. We will send out a call to the community to engage with us as part of this panel and help inform the field. We will also provide them with a way for the community to provide feedback on these provocations through a survey, as well as suggest additional provocations. The co-chairs will analyze the data from the survey to identify questions for our panel as well as relevant themes that can be shared during the panel. This modality will allow scholars across the community to engage and give their perspectives on these provocations, which may also serve as a form of (public) data collection and feedback for a longer journal article.

During the session

We will begin the live session with a 5-10-minute introduction from the co-chairs, introducing the provocations' background, aims, and format of the session, and each of the participants. We then envision having each author take five minutes to present their provocation as fodder for a discussion with our panel.

For each provocation, the authors will introduce their claim and differing perspectives, and Wise and Worseley will each respond in ways that analyze the trade-offs associated with differing positions. Thus, all four authors will not present one after the other, but rather each author's provocation will be followed by a panel discussion, and then we will move on to the next provocation. We envision each five-minute introduction to a provocation to be followed by a five to 10-minute discussion with the panel, with added time for the audience (in-person and online) to add their thoughts as well.

Our panel will consist of two primary discussants, Wise and Worseley, with Borge as a facilitator. The authors will also be seated near the panel to encourage a broader discussion across all the authors and discussants.

Hybrid management and navigation

We plan to use Zoom or the conference video technology to provide real-time access to the discussion. Sebastian Simon will serve as the hybrid ambassador, managing the Zoom logistics and rooms throughout the session with support from Breneil Malcolm. They will co-moderate the management of online logistics, including troubleshooting audio, video, screen-sharing capabilities, and providing support to presenters and discussants as needed.

International Society of the Learning Sciences

Rise Above State-of-the-Art: Designing New Possibilities of GenAI for Knowledge Building

Lydia Cao (Co-chair), University of Toronto, Ly.cao@utoronto.ca
Dina Soliman (Co-chair), University of Toronto, dinaa.soliman@utoronto.ca
Carol K.K. Chan (Co-chair), University of Hong Kong, ckkchan@hku.hk
Carl Bereiter (Discussant), University of Toronto, carl.bereiter@utoronto.ca
Marlene Scardamalia (Discussant), University of Toronto, marlene.scardamalia@utoronto.ca
Shotaro Naganuma, Kyushu University, naganuma.shotaro.062@m.kyushu-u.ac.jp
Tsubasa Minematsu, Kyushu Institute of Technology minematsu@ai.kyutech.ac.jp
Jun Oshima, RECLS, Shizuoka University, joshima@inf.shizuoka.ac.jp

Jiayu Cheng, Vivian Leung, Yue Zhao, Chen Wang, Bodong Chen
jiayuc@upenn.edu, vyleung@upenn.edu, janie.zhaoy@gmail.com, cbd@upenn.edu
University of Pennsylvania

Thérèse Laferrière, Université Laval, Québec, Therese.Laferriere@fse.ulaval.ca
Stéphane Allaire, Université du Québec à Chicoutimi, sallaire@uqac.ca
Adiel Benjamin, University of Haifa, adiel.benjamin@edtech.haifa.ac.il
Yotam Hod, University of Haifa, yotamhod24@gmail.com

Jianwei Zhang, Yanqing Sun, Sudarshan Kammu, Mei-Hwa Chen
jzhang1@albany.edu, yanqing.sun@sunyempire.edu, skammu@albany.edu
State University of New York

Thomas Underwood, University of Hong Kong, tunderwd@hku.hk

Seng Chee Tan, Yee Yin Tan, Chew Lee Teo, Guangji Yuan
sengchee.tan@nie.edu.sg, yeeyin.tan@nie.edu.sg, chewlee.teo@nie.edu.sg, guangji.yuan@nie.edu.sg
National Institute of Education, Nanyang Technological University

Xueqi Feng, Southern University of Science and Technology, fengxq@sustech.edu.cn

Abstract: This symposium presents cutting-edge research that explores how Generative Artificial Intelligence (GenAI) can be leveraged to enrich human-AI collaboration in CSCL for Knowledge Building (KB). It challenges prevailing educational paradigms of 'learning from AI' and instead it foregrounds collaborative creation, emphasizing how GenAI can be harnessed to enrich epistemic agency empowering learners in knowledge building. Building on four decades of KB research, this symposium aims to enrich both the conceptualization and design of human-AI collaboration for knowledge advancement. Seven interrelated contributions demonstrate emergent KB-AI design approaches, including dialogic approaches that engage learners in collaborative dialogue with AI, and integrated approaches that embed AI affordances into knowledge building environments. These designs span diverse contexts from doctoral interdisciplinary research to K-12 classrooms. Grounded in a design-based approach, this symposium positions these contributions as *conceptual artifacts* for design advancement and envisions new KB-AI synergies that enable emerging forms of CSCL and knowledge creation.

Background and rationale

Since the public release of ChatGPT in late 2022, GenAI has been increasingly debated and adopted across educational contexts. Definitions typically emphasize GenAI's capacity to produce synthetic content via generative modeling (Peñalvo & Ingelmo, 2023), but the educational question is not merely *what* it generates; it is *how* we conceptualize GenAI in relation to learning and the advancement of knowledge. Tan (2024) conceptualizes GenAI in education into three areas: learning *about* AI (GenAI literacy), learning *from* GenAI (GenAI as tutor), and learning *with* GenAI (use of GenAI as partner). While current educational research and practice has focused on learning *from* GenAI, such as a personal tutor and students' personalized learning (e.g.,

 © ISLS

Baillifard et al., 2025), there are relatively less studies which explore on learning *with* GenAI focusing on collaboration (Zheng et al., 2024), and even fewer studies on knowledge building with GenAI (Chen et al., 2023; Lee et al., 2023). This symposium focuses on CSCL, examining the role of GenAI in idea improvement and knowledge creation within knowledge building environments.

Early work by Scardamalia and Bereiter (1987) showed that expert writers engage in *knowledge transforming*—an iterative process of refining and restructuring ideas to achieve coherence and generate new understanding. Writing, in this sense, becomes a means of *knowledge creation* rather than simple communication, a view upon which CSCL and KB research on collective inquiry and idea improvement is built. Extending this to the collective level, KB conceptualizes knowledge building as *idea improvement* through shared cognitive responsibility occurring through idea-centered knowledge building discourse (Scardamalia & Bereiter, 2014). Recent scholarship on essays in a time of GenAI echoes these foundations, positioning writing as an *epistemic act* through which students build, test, and refine ideas (Corbin et al., 2025).

Emerging CSCL research investigates GenAI as a *collaborative agent* that supports co-construction of knowledge, reflection, and shared inquiry. Cress and Kimmerle (2023) argue that meaningful learning can occur through *dialogue* between humans and AI systems, where information is mutually exchanged and challenged to drive knowledge advancement. Empirical studies suggest that the integration of GenAI into collaborative digital environments can enhance real-time communication, task regulation, and collective mean-making (Wei et al., 2025). Conceptual work further advances this discussion by proposing frameworks which outline escalating levels of AI agency—from adaptive instrument to peer collaborator—grounded in CSCL work (Yan et al., 2024). Primarily, the dominant narrative of 'learning from AI' is problematic as it reinforces the 'learning as acquisition' and 'mind as a container' paradigms in education (Bereiter, 2002). Researchers have expressed concerns that a narrow focus on AI applications will result in "a type of digital duct tape to hold together an obsolete industrial-era educational system, automating outdated instruction and assessment" (Dede, 2023, p.35). Recent work in CSCL on GenAI has pointed to the importance of designing generative AI for collaborative learning (Anderson et al., 2025; Hernández-Leo et al., 2025). This symposium seeks to build on current work on Generative AI for collaboration to examine how AI can contribute to the cognitive and social dynamics of collaborative knowledge building and knowledge creation.

Theoretical perspective

Knowledge Building (KB) transforms education from a model of knowledge transmission to engaging students directly in how knowledge in the world advances, with students taking collective responsibility for creating and improving ideas of value to their community (Scardamalia & Bereiter, 2006). Knowledge Building moves beyond constructivist approaches by engaging all students as legitimate members of a knowledge-creating community, in which ideas are treated as real, improvable, conceptual artifacts that, through collective work, can be tested, refined and built upon to advance the state of knowledge in the community and in the world (Scardamalia & Bereiter, 2010). A set of KB principles including epistemic agency, improvable ideas, idea diversity, community knowledge is fundamental for guiding KB research and design (Scardamalia, 2002). This principle- based approach is inherently dynamic and demands continuous adaptation to the evolving landscape of knowledge frontiers. Knowledge Builders attempt to work at the cutting edge of knowledge, driving innovation and fostering a culture of collective cognitive responsibility (Scardamalia & Bereiter, 2006).

There is a limited but growing body of literature exploring the intersection of KB and GenAI. For example, Lee et al. (2023) explored the use of GenAI for sustaining student discourse. The authors proposed that GenAI can support Knowledge Building by identifying and explaining authentic problems, summarizing ideas, synthesizing new perspectives, and prompting diverse viewpoints, suggesting improvements, connecting ideas to broader contexts, guiding students in using authoritative sources, and facilitating reflection. A study by Chen et al. (2023) involved interviewing high school students who integrate GenAI into Knowledge Building in a "World of Religions" course. Student interviews showed that GenAI was valued for its efficiency in information search, inspiring new ideas, promoting collaboration, and providing diverse perspectives. However, opinions on its support for creativity were mixed, with some students finding it limited to fact-based information while others appreciated its ability to spark creativity. This suggests the need to examine and develop KB norms and ethics for AI use in these open environments, where deeper inquiry remains important.

While these initial studies show the promise of GenAI in Knowledge Building contexts, there are also significant challenges. One such challenge is the potential compromise of student epistemic agency, defined as "students being positioned with, perceiving, and acting on, opportunities to shape the knowledge building work in their classroom community" (Miller et al., 2017, p. 1058). Since GenAI can generate seemingly convincing learning artifacts, there is a growing concern that students may forge their epistemic agency and use GenAI as a shortcut instead of engaging in the intellectual heavy lifting themselves. Tan (personal communication, 2025)

illustrated this with a "worst case scenario", where the entire knowledge building process becomes artificial, with one student presenting an AI-generated idea, another student using AI to build on that idea, and a third student evaluating the idea using AI. While this scenario highlights potential risks, it also points toward the emergence of new forms of epistemic agency in the age of AI. Chen (2025, p. 2) proposes viewing "AI as epistemic infrastructure" that fundamentally changes "how knowledge is created, validated, and shared in educational contexts." In this conception, AI is not merely an add-on to human action but reshapes what we do and redefines how knowledge is produced. Chen describes AI systems as "active participants in knowledge co-creation" and highlights the need for "how new forms of epistemic agency emerge within human-AI ecologies" (p. 9).

Key questions

This symposium aims to address the key challenge of designing and examining GenAI for collaborative knowledge building. The following three key questions will guide our collective inquiry and idea advancement:

1. How can GenAI tools, systems, infrastructure be designed to foreground epistemic agency and KB principles?
2. What new knowledge building practices become possible through thoughtful design and integration of GenAI? What new forms of epistemic agency within human-AI ecologies emerge from these design spaces?
3. What are the emerging design approaches to leveraging GenAI for Knowledge Building and what new question emerge for forward-looking work?

Table 1 summarizes the 7 contributions, highlighting their focus, key features, context, and design approach. Two key design approaches are evident across the contributions: (1) *dialogic design* that positions AI as a partner that interacts with learners via direct collaborative conversation to advance knowledge, and (2) *Integrated design* which embed AI capabilities within knowledge building environments.

Table 1

Summary of Seven Research Contributions as Conceptual Artifacts

Contribution	Focus	Key Features	Context	Design Approach
ICRTG & RA-GPT Naganuma, Minematsu, & Oshima (Japan)	ICRTG: Theme generator as a brokering device to create interdisciplinary research topics. RA-GPT: an *agent* to promote students' knowledge-building discourse (e.g., high-level inquiry, literature search).	Theme Generation Research Assistance Idea Elaboration AI-Collaborative Reflective Assessment with KBDeX	3-day workshop	Dialogic & integrated design
CraftPad Cheng, Yu, Zhao, Wang & Chen (US)	Flexible workspace with infinite canvas and AI prompts for teachers to design KB lessons	Infinite Canvas KB Prompt Library Iterative Design Flexible Workflow	Teacher learning; Prototype and Planning	Integrated design
CollaBot Laferrière & Allaire (Canada)	Embedded in Knowledge Forum with interactive buttons to support peer assessment and provide individualized feedback	Peer Assessment Individualized Feedback Refine the form and content of their individual contributions	Tested prototypes and refinement	Integrated design
Expert GAI Study Benjamin & Hod (Israel)	Research study examining how experts across disciplines use AI for knowledge building	N/A	Qualitative Study conducted with 10 Experts	Integrated design
G-Ideas Zhang, Sun, & Kammu, Underwood & Chen (US)	AI interaction spaces (threads) where students—individually or as a group--interact with Gen AI agents to engage in reflective conversations and seek assistance.	Collaborative Learning Spaces AI Interaction Assistance in problem finding, ideation, research writing, team planning and reflection.	Collaborative conversations college class	Dialogic and integrated design

KB-LCT Tan, Tan, Teo & Yuan (Singapore)	AI chatbot, situated within the 'HOME of KB' framework, that supports community-based teacher learning and KB lesson design	KB Principles Learning Lesson co-design Reflection Personalized Support Contextual Knowledge Base	Multiple Case Study of three teachers - KB adoption in schools	Dialogic and integrated design
Storytelling GAI Study Feng & Chan (China)	Student and GAI as epistemic collaborators work for scientific storytelling in KB environment	Shared trajectory of scientific story, KB talk, AI prompts; student-AI interaction	Scientific inquiry; primary students	Integrated design

Dialogic designs engage learners in collaborative dialogue with AI. Contributions adopting dialogic designs include G- Ideas (Zhang et al.), which creates parallel spaces for peer discourse and AI interaction, KB-LCT (Tan et al.), which engages teachers in direct dialogue about KB principles and lesson improvement, and RA-GPT (Naganuma et al.), which serves as an agent to promote knowledge building discourse through activities such as information search and illustration creation. These designs position AI as a collaborative partner in knowledge building (Cress & Kimmerle, 2023), focusing on facilitating productive AI-learner interactions that align with KB principles.

Integrated designs embed AI capabilities within KB environments to enrich existing practices. Contributions include CraftPad (Cheng et al.), which supports teachers' meta-design activities and creative agency; CollaBot (Laferrière & Allaire), which provides background analysis and individualized feedback for student written contributions; ICRTG (Naganuma et al.), which functions as a brokering device by generating interdisciplinary themes for students to evaluate and advance; and Storytelling GAI (Feng & Chan) which examines how AI prompts/scaffolds embedded in KB discourse position AI as an epistemic partner for collective knowledge building. Benjamin & Hod's study investigates how experts across disciplines use AI, which provides insights to inform the design of AI for KB. These designs use diverse methods and tools illustrate how human-AI collaboration can be conceptualized and how AI's capabilities can be leveraged to augment knowledge building environments.

Significance of the symposium

This symposium comes at a critical juncture of rapid advancement of GenAI and growing demands for knowledge creation, addressing a gap in current educational research that has largely focused on learning from Generative AI for knowledge acquisition and individual cognition. Situated in current CSCL work on GenAI, this symposium is premised on well-established KB foundational theories and design-based research for enriching ongoing CSCL academic discourse. In an age of information overload and rapid advances in GenAI, education urgently needs to move beyond knowledge transmission toward collaboration and empowerment of learners as creators of knowledge. These contributions build on current work on GenAI for collaborative inquiry (Cress et al., 2023), advancing AI as a partner in creative knowledge work. Contributions are theoretically focused and align with a variety of methodologies and prototypes of AI tools, with practical classroom examples. We showcase the cutting-edge AI-enhanced design prototypes that foreground KB theories and principles across contexts, including higher education, K-12 classrooms, and professional development.

These diverse but coherent designs have significant implications for CSCL theory, pedagogy and technology, suggesting ways to move beyond simply using AI as "digital duct tape" for existing educational approaches. They show how AI can support collaborative inquiry and genuine transformation of classrooms into knowledge creating communities (Dede, 2023). The diversity of perspectives and contexts represented in the symposium reflects the global effort in advancing AI for KB. Together, contributions of this symposium collectively shed light on how AI can be designed and integrated to enhance knowledge building practices and envision new forms of knowledge creation through dynamic and evolving human-AI partnerships.

Generative AI support for forming interdisciplinary teams and enacting disciplinary epistemic agency

Shotaro Naganuma, Tsubasa Minematsu, Jun Oshima

Theoretical background: Interdisciplinary research practices stem from the social need to solve complex problems (e.g., climate change) (e.g., Repko & Szostak, 2021). Despite the growing recognition of the significance of interdisciplinary learning programs (e.g., Warkentien et al., 2022), a productive learning environment for interdisciplinary research practices including team formation and knowledge creation has not fully developed (e.g., Markauskaite et al., 2024). As part of an ongoing design-based research project initiated in 2023 (Naganuma et al., 2024, 2025, 2026), we designed a three-day workshop in 2025 on interdisciplinary research practice. Our

workshop design leverages generative AI (GenAI) in alignment with the theories of boundary crossing, and we have refined our design to implement a knowledge-building design principle: "concurrent, embedded, & transformative assessment" (Scardamalia, 2002).

Context, methods and design of GenAI: Nineteen doctoral students from various departments (e.g., engineering, informatics) participated in a workshop to develop interdisciplinary collaborative research proposals using two GenAI tools: Interdisciplinary Collaborative Research Theme Generator (ICRTG) generated potential research themes as boundary objects, while Research Assistant GPT (RA-GPT) promoted knowledge-creating discourse. The workshop included four activities: (1) promisingness judgment: students in instructor-assigned teams evaluated AI-generated themes twice as exploratory collaboration experiences; (2) team formation: students self-organized final teams through four networking rounds using ICRTG to identify well-matched collaborators; (3) proposal development: teams created research proposals, enhanced in design-based research 2025 round with collaborative reflective assessment -- students visualized their discourse using KBDeX (Oshima et al., 2012) and consulted RA-GPT to identify additional disciplinary ideas; and (4) final presentation: teams delivered 10-minute expert-assessed presentations. Two research questions include: (1) How did GenAI support self-organized interdisciplinary team formation?, and (2) How did GenAI-enhanced collaborative reflective assessment transform their discourse?

Findings and implications: Results revealed that 1) some learners formed their teams when they they judged the AI-suggested themes as promising for the members, while others struggled to find collaborators, and 2) several teams incorporated disciplinary ideas from the GenAI-enhanced collaborative reflective assessment into their discourse. These findings suggest that in community settings where participants are not well acquainted with one another, GenAI can play a valuable role in supporting the early detection of team member fit in interdisciplinary contexts. They also highlight that current GenAI-enhanced collaborative reflective assessment using GenAI and KB analytics might empower learners to mobilize additional disciplinary resources toward exercising their epistemic agency for knowledge advance more productively.

This study highlights two directions for future research on using GAI for interdisciplinary research and knowledge building - The first concerns the development of AI systems that recognize participants' movements from images to visualize real-time interdisciplinary distance within a community to suggest potential interdisciplinary team members, especially for those who struggle to find collaborators. The second concerns whether AI systems that transcribe learners' oral discourse in real time and provide GenAI-enhanced collaborative reflective assessment on demand—at the moment learners require it—can increase the contributions and disciplinary epistemic agency within the team. Implications of development of AI systems and learning-analytics collaborative assessment for enhancing epistemic agency will be discussed.

CraftPad: A Workspace for Teachers to Sketch and Refine Knowledge Building Designs

Jiayu Cheng, Vivian Leung, Yue Zhao, Chen Wang, Bodong Chen

Theoretical background: Designing effective environments for knowledge building represents a sophisticated craft that demands deep understanding of KB principles, careful balancing of structure and emergence, and a design-oriented mindset. While generative AI has gained popularity in supporting teacher lesson planning, current systems—even the most advanced ones—operate within rigid templates that constrain teacher agency (Cheng et al., 2025). These limitations prevent educators from engaging in the skilled, iterative actions essential to the craft of teaching (Chen, 2025). Our work adopts an alternative approach that positions teachers as designers (Ma & Scardamalia, 2022) rather than mere end-users of AI tools. This perspective necessitates "meta- design"—designing for design itself—which empowers educators to configure and adapt tools in service of their professional craft and educational aspirations. By centering teachers' expertise and creative agency, we aim to develop AI-enhanced environments that support the complex, nuanced work of knowledge building design.

Context, methods and design of GenAI: In response to the limitations of current AI systems for teachers, we designed CraftPad as a flexible workspace that centers teachers' expertise and creative agency. CraftPad features: (1) an infinite canvas (similar to Knowledge Forum) providing space for open-ended ideation and flexible organization; (2) an embedded prompt library informed by KB principles to guide various AI generation tasks in support of teacher design; and (3) a flexible workflow enabling teachers to quickly initiate designs, iteratively refine specific components with their professional input, and compile design artifacts (e.g., lesson designs) that incorporate AI-generated content while preserving the primacy of teacher judgment. This approach embodies our meta-design philosophy by supporting teachers as designers rather than mere consumers of AI outputs.

Findings and implications: The study is currently in progress as part of an ongoing design-based research effort. Early design iterations have shown emerging insights into how CraftPad can support teachers' epistemic agency in AI-assisted lesson design. Teachers are beginning to design lessons using KB principles also taking up

more agency not just focusing on the output -- These observations are guiding the next phase of refinement and empirical testing. Education is fundamentally an epistemic endeavor that requires preserving human judgment when introducing AI for epistemic operations such as lesson plan generation (Chen, 2025). Building on our meta-design approach, CraftPad represents a preliminary attempt to support teachers in designing and iteratively refining their ideas. Its core mission aligns with our perspective by re-centering teacher professional expertise while positioning AI in a supportive role that augments rather than replaces teacher design capabilities. We are planning user studies to evaluate CraftPad's viability in practical settings and to further refine our approach to teacher-AI collaboration in knowledge building design.

Empowering knowledge building with generative AI: Balancing affordances and epistemic agency

Thérèse Laferrière, Stéphane Allaire

Theoretical background: Knowledge building is grounded in principles including *Improvable ideas*, *Idea diversity*, and *Epistemic agency* (Bereiter & Scardamalia, 2006; Scardamalia, 2002). Within the field of CSCL, researchers have emphasized the importance of designing tools that not only scaffold knowledge building but also preserve student ownership and responsibility for advancing community knowledge (Chen, 2025). The study situates its work within this theoretical lineage, stressing that AI should empower learners rather than displace their agency.

Design of GenAI-augmented environment 1.

In the first project entitled "L'ÉCRAN: L'Évaluation Collaborative Réussie des Apprentissages par le Numérique", we designed CollaBot to support peer assessment. *Methods.* CollaBot was embedded within Knowledge Forum (KF), presenting itself as five interactive buttons located at the bottom of a note during its composition. Feedback was addressed by assisting students in refining the form and content of their individual responses to other students on KF. The project emphasizes alignment with core knowledge-building principles such as Improvable Ideas, Idea Diversity, and Epistemic Agency. *Findings:* The tool successfully supports the principle of Improvable Ideas by offering targeted feedback that promotes iterative refinement of ideas. Furthermore, CollaBot's embedded and on-demand feedback mechanisms effectively facilitate formative assessment, enabling students to reflect on their contributions and those made by their peers within the same KB discussion thread for identifying areas for improvement. *Implications:* Given the promising findings, there is a need for further development to enhance CollaBot's scaffolding capabilities and ensure it doesn't stifle students' epistemic agency, allowing them to maintain ownership and control over the knowledge-building process.

Design of GenAI-augmented environment 2.

A separate initiative explored a wider range of GenAI applications in a knowledge-building environment within a graduate-level education course. *Method:* Iterative exchanges with students as users of GenAI and on-going verbatim analysis. *Findings:* This initiative identified six key affordances of GenAI for enhancing knowledge-building practices: 1) Support for formalizing complex questions: GenAI can aid students in constructing well-defined and focused research questions to guide their inquiry process; 2) Avenue identification for further research from recognized sources: GenAI can analyze student contributions and propose relevant literature or research directions based on existing knowledge in the field; 3) Idea identification that demonstrates knowledge advancement: GenAI can evaluate student ideas, highlighting those that expand upon or challenge existing knowledge within a knowledge-building community; 4) Support for student self-assessment: GenAI can provide feedback on student contributions, referencing established evaluation criteria, thereby fostering reflection and self-improvement; 5) Support for community work process analysis: GenAI can analyze group interactions and provide insights into group dynamics, collaboration patterns, and potential areas for improvement; and 6) Speech synthesis support: GenAI can effectively summarize key themes and insights emerging from group discussions, providing a concise overview of collective understanding.

Implications: The presentation emphasizes the importance of implementing GenAI in a manner that prioritizes student agency and fosters meaningful learning experiences. The team focused on co-designing GenAI tools and uses in a way that empowers students and promotes their active engagement and collaboration. It also stresses the relevance of incorporating knowledge-building principles in the design and deployment of GenAI tools. We will conclude by acknowledging the nascent, and promising, stage of research on generative AI and knowledge building, emphasizing the need for ongoing investigation into the long-term implications of this integration.

 © ISLS

How do experts use ai to build knowledge? A study of practices across the disciplines

Adiel Benjamin, Yotam Hod

Theoretical background: This research explores how expert knowledge builders use Generative AI (GenAI) as a way to inform and refine established knowledge building design principles (Chen & Hong, 2016; Chan & van Aalst, 2018; Cohen & Hod, 2021; Scardamalia, 2002; Zhang et al., 2011). As GenAI becomes increasingly integral across fields, understanding its role in knowledge creation is crucial for educational researchers aiming to design human-centered, authentic learning environments that address contemporary challenges. By examining expert GenAI practices, this study seeks to align educational strategies with real-world knowledge building approaches, equipping educational systems to meet the demands of an AI-integrated future while emphasizing human elements of learning.

Context and methods: This study addresses two key questions: (1) How do expert knowledge builders intentionally use GenAI to advance collective knowledge in their field? (2) How do these practices align with established knowledge building design principles? Using a grounded theoretical approach (Charmaz, 2008), we conducted semi-structured interviews that were video recorded and screen captured with 7 experts from diverse disciplines, including the humanities, social and natural sciences, and industry. Participants demonstrated their GenAI use, detailing their rationale, tools, and methods employed in knowledge building tasks. For this symposium, we focus on the second research question to advance discussion about the knowledge building principles. From this analysis, six design principles emerged as particularly salient: Improvable Ideas, Epistemic Agency, Collective Responsibility, Democratizing Knowledge, Idea Diversity, and Pervasive Knowledge Building.

Findings: A clear alignment between expert practices and KB principles was the use of GenAI for improvable ideas. Participants iteratively refined their work using GenAI rather than seeking a single, final answer- this iterative process was deeply integrated with epistemic agency, as experts were not passive users; they actively decided when and how to use GenAI, critically evaluated its outputs, and asserted human responsibility. By probing the system's limits and cross-checking information, they affirmed the importance of discernment in AI-augmented work and demonstrated how users must own their inquiry trajectories.

Principles related to community knowledge were also evident, though with some nuance. Collective responsibility emerged in collaborative settings where GenAI acted as a social mediator to stimulate group discussion and shared evaluation, though this practice was less frequent than individual use. The principle of democratizing knowledge was powerfully embodied by a participant with dysgraphia who described GenAI as an enabling technology that allowed him to communicate ideas more fluently, thereby reducing barriers to academic participation. GenAI also supported idea diversity, acting as a "consultant" that could offer alternative perspectives. However, experts noted this required careful prompting, as GenAI could otherwise reinforce dominant narratives, highlighting the need for critical awareness from the user.

Finally, the findings aligned with the principle of pervasive knowledge building, which holds that knowledge creation permeates life across all contexts. Participants frequently drew connections between personal and professional GenAI experimentation. One noted, "when I use it in an everyday way I see that as an opportunity to think about how I can use that in a professional way." This cross-pollination of inquiry demonstrates how GenAI supports knowledge building as a continuous, lived activity, making the tool a shared epistemic resource that enables more equitable and distributed pathways for innovation and discourse.

Implications: This study shows that KB design principles are enacted through an expanded range of practices (Zhang et al., 2011). For example, expert knowledge builders used GenAI not just to improve ideas, but to plan and streamline workflows, automate tasks, and test knowledge through code and simulation. These activities are not auxiliary, but essential in optimizing KB principles to a GenAI-enhanced context. To remain relevant, KB must evolve to acknowledge AI not only as a tool but as a co-actor in the community. Expanding KB principles in light of experts' GenAI practices responds to the call for designing AI practices that serve meaningful human goals (OECD, 2025). As one of the participants explained, "These tools actually make a good researcher even better… but the usage needs to be educated, smart, and appropriate".

Generative Ideas (G-Ideas): Configuring Human-AI Partnerships for Knowledge Building

Jianwei Zhang, Yanqing Sun, Sudarshan Kammu, Thomas Underwood, Mei-Hwa Chen

Theoretical background: While GenAI can be used to enhance student performance in various cognitive tasks such as problem solving and writing, student reliance on external AI support risks shortcutting the effortful

 © ISLS

cognitive processes required for high-quality thinking, causing "cognitive debt" over time (Kosmyna et al., 2025). The Generative Ideas (G-Ideas) project aims to design productive human-AI interaction that fosters generative thinking processes for knowledge building. Students take on high-level agency for advancing ideas through progressive inquiry with GenAI. Our design builds on insights gained from how professional researchers interact with AI agents while working with their team members to solve problems and create knowledge (Anantrasirichai & Bull, 2022; Crimaldi & Leonelli, 2023; Dell'Acqua et al., 2023; Gupta et al., 2023; O'Neill et al., 2022). Knowledge builders work as intellectual stewards to orchestrate human and AI-based processes for continual idea improvement in a dynamically changing context (Zhang, 2026).

Design innovation: Generative Ideas (G-Ideas) was designed to enhance student-driven knowledge building featuring idea-centered inquiry and progressive discourse (Scardamalia & Bereiter, 2006; van Aalst, 2009; Zhang et al., 2022). G-Ideas integrates peer collaboration with AI interaction. It includes (a) peer collaboration spaces for students' collaborative discourse and (b) AI interaction spaces (threads) where students—individually or as a group--interact with GenAI agents. The spaces of student discussion and student-AI interaction are interconnected, with options for selecting/copying information from one space to the other. Information from student discussions is used as context data to inform GenAI processing while the output from GenAI can be used to enrich students' discussions. Students' interactions with GenAI are scaffolded to achieve various epistemic needs, such as to assist problem finding, idea generation and improvement, research writing, and team reflection. Teachers can create and customize scaffolds—organized as workflows and think types—to guide student interaction with AI, prioritizing students' ideas as the focus of the interaction. For example, when taking the choice of idea generation and improvement, G-Ideas first prompts students to generate initial wonderings and tentative ideas and then connects students with the AI agent to improve the ideas while uncovering deeper issues. For transparency and accountability, the histories of student-AI interactions are shared with their teacher. Students can also choose to share the records with peers as a resource to enrich their knowledge work.

Methods: We are testing how students work with G-Ideas to support authentic knowledge building processes through design-based research among high school and college classes. In each context, students work as collaborative teams to conduct research and design to address authentic challenges. Data collection traces student engagement in the research processes supported by their interaction with GenAI. Student learning outcomes are evaluated using teacher-created assessments.

Findings: The study conducted in the college course compared student learning in three conditions: 1) without using AI, 2) using AI without reflective guidance, and 3) using AI with the reflective guidance provided by G-Ideas. Students under condition 2 using AI outperformed those of condition 1 in their project work during the class, but they scored lower in the final exam (conducted without AI help). Their use of AI improved their project assignments but not their understanding of the core concepts. In condition 3, with reflective guidance of AI interaction, students demonstrated the highest performance both in the project work submitted and the final exam (without AI). These results highlight that incorporating a reflective process enhances students' understanding of the underlying concepts related to their design work.

Implications: G-Idea design and research showcase configurations of student-AI partnerships for addressing different epistemic needs in knowledge building. Depending on the roles for GenAI to play in knowledge building, human-AI partnerships may be configured in different ways, ranging from *outsourcing* well-defined work components (e.g., notetaking, artefact generation) to GenAI, to *consulting* GenAI about the knowledge and needs of a field, and *collaborating* with GenAI as a co-inquirer to seek continual idea advancement. As a guiding principle, it is essential to keep student-generated ideas at the center and support students' reflective judgment for co-structuring creative knowledge practices performed by human and AI systems for continual knowledge advancement, personal growth, and community flourishing (Zhang, 2026).

Leveraging large language models as a teacher learning companion for knowledge building

Seng Chee Tan, Yee Yin Tan, Chew Lee Teo, Guangji Yuan

Theoretical background: The design of AI in this study is aligned to Gao et al.'s (2025) concept of human-centred human-AI collaboration (HCHAC) in which AI systems are designed to augment human capabilities while respecting human autonomy and values. Collaboration is framed as a partnership, where humans lead and AI supports. In terms of learning, it is aligned with the theory of social generative AI (Sharples, 2023), where Generative AI (GenAI) chatbot is leveraged as a dialogic partner that a human learner, including teachers, can converse with to co-construct meaning and artefacts, and not simply a prompt-response tool.

Design of the AI innovation: This study involves developing and evaluating a GenAI chatbot, the Knowledge Building Learning Companion for Teachers (KB-LCT), which was developed based on large language models to support teachers in learning about Knowledge Building (KB) and designing KB lessons. Focusing on KB principles (Scardamalia & Bereiter, 2010) and developed on a customisable AI-powered platform, SchoolAI (2023), the KB-LCT seeks to support teachers in three ways: (a) learning the 12 principles of KB, (b) co-designing a lesson based on KB, (c) reflecting over potential implementation problems in their KB lesson design. The chatbot was designed with prompts to guide teacher interactions, featuring personalization such as personalized greetings and context-based answers. To reduce "AI hallucination" and to provide a more contextualised knowledge base, we used the retrieval augmented generation technique (Lewis et al., 2020) by uploading research articles on KB and author-generated documents on common challenges of implementing KB in local contexts. Implementing the KB-LCT was underpinned by a teacher-professional learning community approach. Using the 'HOME of KB' model Human-cOmputer-coMmunity-Exploration of Knowledge Building) (Teo, 2024), teacher participants went through a) individual interactions with the KB-LCT, b) community-level discussions facilitated by a professional learning team leader, a teacher who had experience designing KB with the KB- LCT, whilst c) implementing their lessons.

Methods: Using a multiple case study approach (Stake, 2013), we interviewed three teachers who had learnt to design and implement KB with the KB-LCT and examined their interactions with the KB-LCT over five KB lessons. Guided by the following research questions: "What affordances of the KB- LCT and the HOME of KB do the teachers experience? What limitations do they perceive?", we analysed the data from the semi-structured interviews.

Findings: Our preliminary findings showed that the teachers exercised their agency to adapt suggestions provided by the KB-LCT and made decisions based on their teaching contexts and knowledge gained from past professional development in KB. Regarding the affordances, the teachers commented that the KB-LCT acted as a learning companion that is a) user-centric, b) contextualized, c) personalized, d) useful in triggering reflections, and e) capable of providing pedagogical suggestions aligned with KB. Despite some limitations, such as the need for prerequisite competencies and expanded features, this study tested a preliminary application of large-scale language models in teacher training for knowledge building, marking it as one of the pioneering works in this emerging field.

Implications for AI/ knowledge building/CSCL: This study provides the empirical evidence of how HCHAC system can be designed. The design of the HOME of Knowledge Building (KB) environment empowers teacher participants to exercise their professional agency by actively seeking information through a GenAI chatbot (AI-in-the-loop). This process is enriched through peer interactions, which foster deeper engagement with knowledge building principles and support the enactment of these practices in classroom settings.

Fostering Scientific Storytelling through Student–AI Knowledge Building Collaboration in Primary Science

Xueqi Feng, Carol Chan

Theoretical background: Scientific storytelling has been conceptualized as an epistemic practice that supports learners in organizing knowledge, constructing causal explanations, and narrating inquiry through disciplinary reasoning (Avraamidou & Osborne, 2009). Grounded in models of how scientific communities advance knowledge, KB emphasizes progressive discourse and shared epistemic agency, where learners collaboratively pose questions, improve ideas, and construct explanations (Bereiter et al., 1997). As an emerging epistemic partner, GenAI can engage students in dialogic interactions that prompt elaboration, raise epistemic questions, and encourage reflection (Cress & Kimmerle, 2023). In this case, GenAI amplifies learners' epistemic agency, contributing as a co-participant in KB discourse.

Design: This study implemented a KB environment in which a GenAI scaffolds were embedded as an epistemic collaborator to support scientific storytelling in primary science. Students co-constructed an epistemic trajectory—identifying phenomena, formulating problems, building explanations, seeking evidence, and engaging in theoretical discourse—and engaged in sustained inquiry on Knowledge Forum. GenAI played a dynamic role across phases: supporting conceptual exploration during collaborative discourse, assisting experimental design in hands-on investigations, and facilitating narrative refinement through reflection. In the final phase, students co-developed criteria for high-quality scientific stories and used GenAI to iteratively improve their work.

Methods: Forty-eight Grade 5 students engaged in the topic of Earth Science, collaboratively investigating natural phenomena including earthquakes, tsunamis, and volcanic eruptions. Data sources included

 © ISLS

pre- and post-scientific stories, domain understanding tests, student Knowledge Forum notes and AI interactions, and classroom artifacts.

Findings: Quantitative analyses revealed significant improvements in students' storytelling quality, domain understanding, and the depth of their KB discourse. Analyses indicated that the quality of students' contributions in KB and AI dialogues significantly predicted storytelling performance, which in turn predicted gains in domain understanding. These findings suggest a mediating role of storytelling in linking collaborative discourse with domain learning. Qualitative analysis identified four themes that supported the emergence of scientific storytelling including epistemic co-design, dialogic augmentation, evidence grounding, and discursive refinement.

Implications: This study introduces a generative model of human–AI collaboration that positions AI as an epistemic collaborator in student-driven inquiry. The findings contribute to the design of AI-mediated, inquiry-based learning environments that promote deeper scientific understanding and epistemic agency through scientific storytelling.

Symposium structure

This symposium uses a structured poster session format designed to foster collective knowledge advancement and interactive engagement with seven innovative KB-AI designs. Following introduction on problem framing (5 minutes) and short presentation of studies (3 minutes each), participants will move freely among posters (supported with artefacts and tools on notebooks) engaging with research teams in conversations about their design innovations and they can post their views on poster boards / online platforms (40 minutes). The symposium will conclude with the commentary from discussants and dialogue among participants (20 minutes). Rather than treating these as individual research projects, we invite participants to approach them as interconnected improvable conceptual artifacts – they are encouraged to interact and engage in *meta-design discourse* to rise above existing design approaches and envision new KB-AI synergies that not only enhance existing KB practices but also enable new forms of knowledge creation. Participants are encouraged to consider these questions:

1) What insights emerge when you consider these artifacts as a whole?
2) What is missing when you look at the collection? What new design spaces open up?
3) What new KB-AI possibilities become visible when you consider any single approach or rise-above combining different approaches?
4) What new forms of epistemic agency within human-AI ecologies emerge from the AI-KB design space?

The session concludes with commentary from Dr. Marlene Scardamalia and Dr. Carl Bereiter, who pioneered Knowledge Forum for CSCL and conceptualized Knowledge Building as knowledge creation advancing the idea of students as knowledge creators, which now becomes more important in this GenAI age. Their reflections will illuminate how these emerging AI designs relate to core KB visions enriching CSCL while opening possibilities for new conceptualization and innovative designs transforming human-AI collaboration. We envision the dialogue will inspire collective advances advancing new KB-AI synergies for enabling new forms of CSCL and knowledge creation.

References

Anantrasirichai, N., & Bull, D. (2022). Artificial intelligence in the creative industries: A review. Artificial Intelligence Review, 55(1), 589–656. https://doi.org/10.1007/s10462-021-10039-7

Anderson, E., Lin, G. C., Farid, A., Fenech, M., Hanks, B., Klopfer, E., Doherty, E., Hirshfield, L., Ko, M. M., Foltz, P., Nguyen, H., Nguyen, V., Ludovise, S., Santagata, R., Cao, L., Scardamalia, M., Soliman, D., Resendes, M., Khanlari, A., Costa, S., Chan, C. K., & Nguyen, A. (2025). Exploring GenAI Technologies Within Collaborative Learning. In Oshima, J., Chen, B., Vogel, F., & Järvelä, J. (Eds.), Proceedings of the 18th International Conference on Computer-Supported Collaborative Learning - CSCL 2025 (pp. 499-506). International Society of the Learning Sciences.

Avraamidou, L., & Osborne, J. (2009). The role of narrative in communicating science. *International Journal of Science Education*, *31*(12), 1683-1707.

Baillifard, A., Gabella, M., Lavenex, P. B., & Martarelli, C. S. (2025). Effective learning with a personal AI tutor: A case study. *Education and Information Technologies*, *30*(1), 297–312.

Bereiter, C., Scardamalia, M., Cassells, C., & Hewitt, J. (1997). Postmodernism, knowledge building, and elementary science. *The Elementary School Journal*, *97*(4), 329-340.

Cao, L., Scardamalia, M., Bereiter, C., Costa, S., Resendes, M., Soliman, D., Khanlari, A., Steipe, B., Zhang, J., Chen, M.-H., Chen, B., Laferrière, T., Yang, Y., Chan, C. K. K., Zhu, G., Teo, C. L., Castillo, F., Raman, P., & Massey, L. (2024). Knowledge Building and Generative AI to Advance Community Knowledge.

General Proceedings of the 4th Annual Meeting of the International Society of the Learning Sciences 2024, 106–109.
Chan, C. K. & van Aalst, J. (2018). Knowledge building: Theory, design, and analysis. In F. Fischer, S. R. Goldman, C. E. Hmelo-Silver, & P. Reimann (Eds.), *International Handbook of the Learning Sciences* (pp. 295-307). Routledge.
Chan, J., Akamatsu, M., Vargas, D., Kawerau, L., & Gartner, M. (2024). Steps Towards an Infrastructure for Scholarly Synthesis (arXiv:2407.20666). *arXiv*. https://doi.org/10.48550/arXiv.2407.20666
Charmaz, K. (2008). Grounded Theory as an Emergent Method. In S. N. Hesse-Biber & P. Leavy (Eds), *Handbook of Emergent Methods* (pp. 155- 172). The Guilford Press.
Chen, B. (2025). *Beyond Tools: Generative AI as Epistemic Infrastructure in Education* (arXiv:2504.06928). arXiv. https://doi.org/10.48550/arXiv.2504.06928
Chen, B., & Hong, H. Y. (2016). Schools as knowledge-building organizations: Thirty years of design research. *Educational Psychologist*, *51*(2), 266–288.
Chen, B., Scardamalia, M., & Bereiter, C. (2015). Advancing knowledge-building discourse through judgments of promising ideas. *International Journal of Computer-Supported Collaborative Learning*, *10*(4), 345–366. https://doi.org/10.1007/s11412-015-9225-z
Chen, B., Zhu, X., & Díaz del Castillo H., F. (2023). Integrating generative AI in knowledge building. *Computers and Education. Artificial Intelligence*, *5*, 100184-. https://doi.org/10.1016/j.caeai.2023.100184
Cheng, J., Wang, C., & Chen, B. (2025). Evaluating Student Agency in AI-Generated Lesson Plans. In the Proceedings of the ICLS Conference. ISLS: Helsinki, Finland.
Cohen, E., & Hod, Y. (2021). Enriching the informing cycle of Knowledge Building Communities by investigating students' interpretations of design principles. *Interactive Learning Environments*, *31*(7), 4382–4394.
Cress, U., & Kimmerle, J. (2023). Co-constructing knowledge with generative AI tools: Reflections from a CSCL perspective. *International Journal of Computer-Supported Collaborative Learning*, *18*(4), 607– 614.
Crimaldi, F., & Leonelli, M. (2023). AI and the creative realm: A short review of current and future applications. https://doi.org/10.48550/arXiv.2306.01795
Dede, C. (2023, August 6). What is Academic Integrity in the Era of Generative Artificial intelligence? *Silver Lining for Learning*. https://silverliningforlearning.org/what-is-academic-integrity-in-the-era-of-generative-artificial-intelligence/
Dell'Acqua, F., McFowland, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Krayer, L., Candelon, F., & Lakhani, K. R. (2023). Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality (SSRN Scholarly Paper 4573321). https://doi.org/10.2139/ssrn.4573321
Gao, Q., Xu, W., Pan, H., Shen, M., & Gao, Z. (2025). Human-centered human-AI collaboration (HCHAC). *arXiv preprint arXiv:2505.22477*. https://doi.org/10.48550/arXiv.2505.22477
García Peñalvo, F. J., & Vázquez Ingelmo, A. (2023). What Do We Mean by GenAI? A Systematic Mapping of The Evolution, Trends, and Techniques Involved in Generative AI. *IJIMAI*, *8*(4), 7–16.
Gupta P, Nguyen TN, Gonzalez C, Woolley AW: Fostering collective intelligence in human–AI collaboration: laying the groundwork for COHUMAIN. Top Cogn Sci 2023. https://doi.org/ 10.1111/tops.12679
Hernández-Leo, D., Theophilou, E., Oshima, J., Minematsu, T., Matsueda, K., Naganuma, S., Cress, U., Chen, B., Zhu, X., del Castillo, F. D., Chen, W., Lyu, Q., Chen, X., Zheng, L., & Weinberger, A. (2025). Collaborating With Generative Ai for Learning?. In Oshima, J., Chen, B., Vogel, F., & Järvelä, J. (Eds.), Proceedings of the 18th International Conference on Computer-Supported Collaborative Learning - CSCL 2025 (pp. 525-533). International Society of the Learning Sciences.
Kosmyna, N., Hauptmann, E., Yuan, Y. T., Situ, J., Liao, X. H., Beresnitzky, A. V., Braunstein, I., & Maes, P. (2025). Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task. arXiv preprint arXiv:2506.08872. https://doi.org/10.48550/arXiv.2506.08872
Lee, A. V. Y., & Tan, S. C. (2017). Promising Ideas for Collective Advancement of Communal Knowledge Using Temporal Analytics and Cluster Analysis. *Journal of Learning Analytics*, *4*(3), 76–101. https://doi.org/10.18608/jla.2017.43.5
Lee, A. V. Y., Tan, S. C., & Teo, C. L. (2023). Designs and practices using generative AI for sustainable student discourse and knowledge creation. *Smart Learning Environments*, *10*(1), 59–18. https://doi.org/10.1186/s40561-023-00279-1
Lee, H.-K. (2022). Rethinking creativity: Creative industries, AI and everyday creativity. *Media, Culture & Society*, *44*(3), 601–612. https://doi.org/10.1177/01634437221077009
Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. In *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)*

Ma, L., & Scardamalia, M. (2022). Teachers as Designers in Knowledge Building Innovation Networks. In *The Learning Sciences in Conversation*. Routledge.

Markauskaite, L., Schwarz, B., Damșa, C., & Muukkonen, H. (2024). Beyond disciplinary engagement: Researching the ecologies of interdisciplinary learning. Journal of the Learning Sciences, 33(2), 213–241.

Miller, E., Manz, E., Russ, R., Stroupe, D., & Berland, L. (2018). Addressing the epistemic elephant in the room: Epistemic agency and the next generation science standards. *Journal of Research in Science Teaching*, 55(7), 1053–1075. https://doi.org/10.1002/tea.21459

Naganuma S., Minematsu T., Matsueda K., & Oshima J. (2024). Use of generative AI for boundary crossing in interdisciplinary collaborative research, *CSCL 2024 Proceedings*, pp.67-74.

Naganuma S., Minematsu T., Matsueda K., & Oshima J. (2025). Generative AI-supported collaboration in interdisciplinary research practice, *CSCL 2025 Proceedings*, pp.340-344.

Naganuma, S., Minematsu, T., Matsueda, K., & Oshima, J. (2026). Doctoral students' interdisciplinary research proposal development supported by generative AI. *Education and Information Technologies*, Advance online publication. https://doi.org/10.1007/s10639-026-13952-y

OECD (2025). Empowering learners for the age of AI: An AI literacy framework for primary and secondary education (Review draft). *OECD*. https://ailiteracyframework.org

O'Neill, T., McNeese, N., Barron, A., & Schelble, B. (2022). Human–Autonomy Teaming: A Review and Analysis of the Empirical Literature. *Human Factors*, *64*(5), 904–938. https://doi.org/10.1177/0018720820960865

Oshima, J., Oshima, R., & Matsuzawa, Y. (2012). Knowledge building discourse explorer: A social network analysis application for knowledge building discourse. Educational Technology Research and Development, 60(5), 903–921.

Repko, A. F., & Szostak, R. (2021). Interdisciplinary research process and theory (fourth edition). SAGE Publications, Inc.

Scardamalia, M. (2002). Collective cognitive responsibility for the advancement of knowledge. *Liberal Education in a Knowledge Society*, *97*, 67-98.

Scardamalia, M., & Bereiter, C. (2006). Knowledge building: Theory, pedagogy, and technology. In K. Sawyer (Ed.), *Cambridge Handbook of the Learning Sciences* (pp. 97–118). Cambridge University Press.

Scardamalia, M., & Bereiter, C. (2010). A Brief History of Knowledge Building. *Canadian Journal of Learning and Technology*, *36*(1), 1-. https://doi.org/10.21432/T2859M

SchoolAI. (2023). SchoolAI [Educational platform]. https://schoolai.com/

Sharples, M. (2023). Towards social generative AI for education: Theory, practices and ethics. *Learning: Research and Practice*, 9(2), 159–167. https://doi.org/10.1080/23735082.2023.2261131

Si, C., Yang, D., & Hashimoto, T. (2024). Can LLMs Generate Novel Research Ideas? A Large-Scale Human Study with 100+ NLP Researchers (arXiv:2409.04109). *arXiv*. http://arxiv.org/abs/2409.04109

Stake, R. E. (2013). *Multiple Case Study Analysis*. Guilford Publications.

Tan, S. C., Chen, W., & Chua, B. L. (2023). Leveraging generative artificial intelligence based on large language models for collaborative learning. *Learning: Research and Practice*, 9(2), 125-134.

Tan, S. C. (2024). *Learning from, learning with, learning about, and learning beyond AI*. https://dr.ntu.edu.sg/handle/10356/181097

Teo, C. L. (2024, May 29). Human-cOmputer-coMmunity-Exploration of KB (HOME of KB). Gen AI for educators' professional growth. Workshop at the *Redesigning Pedagogies International Conference*, Singapore.

van Aalst J. (2009). Distinguishing knowledge-sharing, knowledge-construction, and knowledge-creation discourses. International Journal of Computer-Supported Collaborative Learning, 4(3), 259–287. https://doi.org/10.1007/s11412-009-9069-5

Warkentien, S., Goeking, J., Dilig, R., Knapp, L., & Stanley, R. (2022). Interdisciplinary education literature review and landscape analysis. Lucas Education Research, RTI International.

Zhang, J. (2026). Intellectual Stewardship: Re-adapting Human Minds for Creative Knowledge Work in the Age of AI. https://doi.org/10.48550/arXiv.2603.18117

Zhang, J., Hong, H.-Y., Scardamalia, M., Teo, C. L., & Morley, E. a. (2011). Sustaining knowledge building as a principle-based innovation at an elementary school. *Journal of the Learning Sciences*, *20*(2), 262–307.

Zhang, J., Tian, Y., Yuan, G., & Tao, D. (2022). Epistemic agency for costructuring expansive knowledge-building practices. Science Education (Salem, Mass.), 106(4), 890–923. https://doi.org/10.1002/sce.21717

Zheng, L., Fan, Y., Gao, L., Huang, Z., Chen, B., & Long, M. (2024). Using AI-empowered assessments and personalized recommendations to promote online collaborative learning performance. *Journal of Research on Technology in Education*, 1–27. https://doi.org/10.1080/15391523.2024.2304066

International Society of the Learning Sciences

Towards a Robust Theory Toolbox for CSCL: Mechanisms, Models, and AI

Sten Ludvigsen, (co-organizer) University of Oslo, s.r.ludvigsen@iped.uio.no
Davinia Hernández-Leo, Universitat Pompeu Fabra, davinia.hernandez-leo@upf.edu
Peter Reimann, (co-organizer) University of Sydney, peter.reimann@sydney.edu.au
Alyssa Wise, Vanderbilt University, alyssa.wise@vanderbilt.edu
Fanjie Li, Vanderbilt University, fanjie.li@vanderbilt.edu
Hans C. Arnseth, University of Oslo, h.c.arnseth@iped.uio.no
Kenneth Silseth, University of Oslo, kenneth.silseth@iped.uio.no
Rolf Steier, OsloMet, rolste@oslomet.no
Jo Inge J. Froytlog, jo.i.froytlog@usn.no
Michael Baker, (co-organizer, chair and discussant) Télécom Paris, michael.baker@telecom-paris.fr

Abstract: This symposium aims to foster new scientific discussions about how the CSCL (Computer-Supported Collaborative Learning) field can advance towards more robust theories. Such advancements depend on further development and formalization of models, causal mechanisms, and the creative application of different types of generative AI. The problem description is rooted in the increasing gap between theoretical positions and actual data, also emphasizing that data alone cannot explain any phenomenon. The overarching argument is that computational modeling approaches can be employed to formalize models of collaborative scripts, temporal collaborative processes, multi-level collaboration, and multimodal activities and that varied uses of AI can amplify such efforts. These models can be based on causal mechanisms and can be generalized to collaborative learning phenomena rather than just specific populations. We provide a rich view of the CSCL field that includes pathways to theoretical robustness, practical tools and solutions for collborative learning in the age of AI.

Introduction

We observe a widening gap in CSCL — as in other behavioural research — between theories and data. Progress in instrumentation has enabled the capture of collaboration of unprecedented width (duration, number of participants), breadth (multi-level and -modal), and depth (granularity) (Reimann, 2021). The availability of multimodal data with high temporal granularity leads to the question of how this can inform theories of collaboration and learning. The problem is that most of the (psychological, for instance) theories we use in CSCL have not been conceived with this kind of data in mind. They do not predict or explain unaggregated behavioural data such as event sequences, only changes in aggregated variables. They do not account for the situatedness of cognition and interaction, but abstract away from context. They are (meant to be) tested with statistical models for linear relations that cannot account for feedback relations in dynamic and complex systems. Hence, there is a need to develop theory and along with that, to reflect on what kinds of theories the field needs. This involves, looking into neighboring disciplines for considering epistemic partnering.

The conference motto, "Partnering with Purpose," is interpreted epistemically here: We ask which concepts, theories, and methods from related fields should inform multidisciplinary CSCL research. This question is significant because there is a consensus against a single, unified theory of CSCL and a standardized methodology (Wise & Schwartz, 2017; Rummel, 2018). We employ the metaphor of the *toolbox* to illustrate this diversity; the items inside are varied but useful. While others have proposed what should be in a social sciences theory toolbox more generally (e.g., Bell et al., 2012), this symposium focuses on CSCL. We suggest including a specific type of theory tool—namely, *mechanisms*—and provide examples of such. (There will be other types of tools in the box.) The CSCL researcher also picks tools from a *method toolbox*, and we recommend additions to the method box: computational modeling and application of generative AI. We suggest these types of tools because they can be instrumental in developing robust theories and methods. Identification of collaborative learning mechanisms, computational modelling, and causal explanations help ensure that CSCL enhances its collective and cumulative capacities to develop as an interdisciplinary scientific field.

Over the last 25 years, philosophers of science have studied how descriptions of mechanisms and mechanical explanations address causality in fields such as biology, neuroscience, cognitive sciences and cognitive psychology (Bechtel, 2009; Hedström & Ylikoski, 2010; Ludvigsen and Frøytlog, 2026). The argument here is that mechanistic explanations, as a meta-perspective, may improve the quality of learning research because it offers a pathway to more precise explanatory models that can be used across research communities in learning

 © ISLS

sciences and CSCL. Mechanical explanation can give us deep insight in a central concern in CSCL; the multiple levels and timescales problem in the study of collaborative learning (Ludvigsen, et al 2025). A mechanism-based explanation "...provides an answer to a how-question underlying a causal why-question. In other words, a mechanism-based explanation tells us how the cause produces its effects by describing the process by which this happens." (Ylikoski, 2018, p. 16). In the social and behavioral sciences, causal mechanisms are used to explain specific outcomes (observed in a case study or experiment, for example) and to develop theories about (social and psychological) mechanisms (Elster, 2015; Frith & Frith, 2012). Following Ylikoski (2018), we call the former a *causal scenario* and the latter a *mechanism scheme*. An example of a mechanism scheme from sociology is Schelling's (1971) segregation mechanism, and a corresponding causal scenario is a specific application of the mechanism to explain an empirical case (e.g., Caetano & Maheshri, 2017). While a single mechanism scheme can be fairly compact, applied to hypothetical or empirical scenarios, it often needs a combination of mechanisms, all of which interact with situational context over time. Hence, simulating how the mechanisms unfold in a specific scenario requires the use of computers to achieve inferential robustness.

Computer modelling of processes takes place in many sciences, but we see very few attempts to do this in CSCL, for one exception see Abrahamson et al. (2007). Modelling needs a strong conceptual base both with regards to theory and methodological stances (Guest & Martin, 2021). Computer modelling is a kind o formalization strategy has been a success in science and engineering, where mathematics is the unifying language. Other forms of formalization are logics (Bringsjord, 2008) and specifications for programming a computer (Cooper & Guest, 2014). Whilst very rare in CSCL and the Learning Sciences in general, formalization of theory (for example, in form of a computer model) is frequently practiced in other collaboration sciences (for instance: Cushman & Gershman, 2019; van Eijndhoven et al., 2023). We treat computational modelling as a form of theory building, different from theory testing through statistical or computational data modelling (see also Wise & Schwarz, 2017). Epistemically, we see the key function of computational models in helping to answer 'what-if' questions: what if this were the case rather than that? In other words, models have an inferential function; they support counterfactual reasoning (Woodward, 2003; Ylikoski & Aydinonat, 2014).

Based on mechanistic models, one can conceptualize a set of tools to explain collaborative learning at multiple levels. These tools could include mapping processes from behavioral data (clicks) to constructs used in developing analytic models and formalized models. Additionally, we need to create new tools for explaining learning as temporal processes. While this has been addressed in numerous qualitative studies for some time, new techniques allow for analyzing richer datasets over time based on formal models. This approach is connected to the multi-level nature of collaboration, as it moves from neuro-level to cognitive and social processes or examines learning in institutional settings. Gen-AI tools are already being used for multi-model coding (Cohn et al., 2025), conceptual refinement, and theoretical exploration (Wise, 2024). Core concepts in CSCL, such as scripts, orchestration, and many practical complexities, can now be shifted from design issues toward mechanism schemas involving modeling and formalized approaches. This transition enables us to move from mere descriptions to robust explanations. Agent-based models (ABMs), Bayesian network models, and cognitive architecture–based models (such as those based on ACT-R) each offer complementary ways of understanding group learning phenomena. ABMs simulate learning at the collective level by representing individuals as autonomous agents who interact according to simple behavioural rules (Wilensky & Rand, 2015). Through these interactions, emergent group-level patterns can be studied, such as how peer influence, communication structures, or heterogeneous strategies affect overall learning outcomes. New work leveraging stochastic generative AI models with memory state tracking as agents offers particualrly exciting promise here. Bayesian network models emphasise the probabilistic dependencies among variables. They can capture how beliefs and knowledge states develop within a group, making them well-suited for modelling shared reasoning processes and the propagation of misconceptions or consensus formation. Models based on cognitive architecture, like ACT-R, add a psychologically grounded layer by simulating the cognitive mechanisms of individuals, such as memory retrieval, attention, and problem-solving strategies. When extended to group contexts, these models can show how individual cognition interacts with social processes—e.g., how shared tasks are coordinated or how cognitive load influences collaboration. Taken together, these modeling approaches provide a multi-scale perspective: from micro-level cognition to meso-level interactions to macro-level emergent learning dynamics.

The issues highlighted here, and more generally the challenges of knowledge integration across studies with different epistemologies within learning science and CSCL research, should encourage the community to engage in discussions about future studies and conceptual work in CSCL. Additionally, the wave related to AI/GenAI affects the CSCL field and should inspire us to reconceptualize the conditions for human learning (Baker and Reimann, 2024) as well as how to define an appropriate unit of analysis. To achieve this, both modeling and mechanical explanations could contribute steps toward strengthening CSCL theory. Research questions that will be addressed in and across the papers and presentations are:Can theoretical modelling foster a closer

connection between qualitative and quantitative research? New kinds of qualitative research? Can it help bridging if not unifying different epistemic cultures in CSCL?

- How can mechanism in collaborative learning be described and how can mechanical explanation contribute to formalizations of theory?
- What is the role of formalisation and computer simulation in theoretical modelling? When are these useful and when are they not in CSCL? For example, does it assist in conceptually integrating computer science and AI into CSCL?

Since scientific fields are assessed based on their conceptual development and their ability to describe, understand, and explain the phenomena that constitute the field, we will discuss what makes a theory solid and robust in CSCL. In Reimann's contribution, he emphasizes that we, as a community, need to develop robust generative models that can connect theory and data. He argues that we must use advanced statistical techniques to test theories. Wise and Li advocates for advancing tools based on sophisticated computational analysis and modeling techniques. She also proposes three hypotheses for how we can use generative AI to improve coding, refine concepts, and test theories. Hernández-Leo argues based on historically important CSCL contribution such as scrips and orchestration that we can improve the research by develoing insight into mechanisms and formalized models. Based this approach, we can develop formalized models, structural analyses, and address complexities in practice along with epistemic challenges. This could contribute to the development of theoretical robustness and practical problem-solving in educational settings. Ludvigsen et al argues for a meta-stance that can advance the CSCL field. This meta-stance should be built on a more unified conceptual vocabulary. Here, it is necessary to include descriptions and analysis of mechanisms at different analytical levels and mechanistic explanations that address causality.

Paper 1: Generative models as bridges between theory and data

Peter Reimann

I propose computational modelling, which yields models that produce data, for the CSCL theory and method toolbox. As a methodological tool, models are needed to deal with multimodal, multilevel, high-density temporal data, now ubiquitous in CSCL research. Also, the epistemological role of models is important to consider because it has implications for the nature of theories.

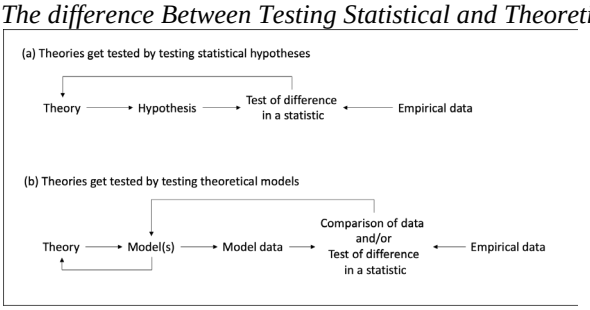
Computational modelling for data analysis

CMs are particularly needed for learning analytics in the context of collaborative learning, sometimes called collaboration analytics (Martinez-Maldonado et al., 2021). Not only can interactions between students collaborating via computer be traced in form of log files, but using ambient and sensor technologies we can now trace interactions that take place in physical space as well, ranging from movement and gesture to eye gaze and bio signals. The wide availability of multimodal data of high temporal granularity in combination with rapidly advancing data analysis methods leads to the question of how these can inform theories of collaboration and learning.

Phrasing the question in this way might seem at putting the cart before the horse: Shouldn't it be the theories and research questions that drive the development of methods? This is a chicken-egg question of little practical relevance: Methods/instruments have always influenced what questions are raised and which explanations are suggested (Latour & Woolgar, 1986). Somewhat ironically perhaps, we need more theory because we have more and different data now, data of different kind at different levels, and more streaming data. This creates challenges for statistics and the ways we use statistics (Franke et al., 2016). For instance, measures of significance that depend on the number of observations quickly become useless as criteria for selecting important variables. Machine learning offers many other methods for identifying potentially relevant relations in large sets of data. However, a purely inductive, theory-agnostic approach is ultimately not satisfying, certainly not for advancing scientific theories. Those need to be tested more directly. A problem we face with most collaboration theories is, however, that they have not been conceived with these new kinds of data in mind.

Since the term “model” is used in many ways (for examples see Magnani and Bertolotti, 2017), Figure 1 has the purpose to make clearer the distinction between the two interpretations of relevance here: statistical models—i.e., models of data—and theoretical models ((a) and (b) in the figure). In modern science, models are predominantly computational in nature. These computer simulations are generative: they produce model data that can be directly compared and/or via statistical tests with empirical data. Typical examples are climate models, models of disease spread, and models of school segregation. Theories are usually too general and too abstract to be compared directly with (statistical models of) empirical data.

Figure 1
The difference Between Testing Statistical and Theoretical Models



Since statements in theories are designed to address general cases, they tend to abstract away from individual and contextual differences. In CSCL, however, these differences are essential for explaining relevant phenomena; for instance, collaboration and learning are considered mediated, if not constituted, by technologies (a type of 'context'). And both learners and context change over time; otherwise, we'd not be studying learning. Thus, the kinds of models useful for CSCL research need to provide parameters to account for differences if not for individual learners (or groups), then for archetypical learners (or groups), and they need to generate event sequences (minimally: summary statistics of those). This means models need to be executable implemented as computer programs.

Up to now, computer models that simulate collaborative learning are much less prevalent than statistical and inductive models. This is, for example, evidenced by the absence of a single instance in the comprehensive overview of collaboration analytics methods by Martinez-Maldonado et al. (2021). Simultaneously, there is a significant amount of model-based research in other areas of collaboration research, such as on teams in organizational science (Blanco-Fernández et al., 2024). Differences in methods education between the disciplines may explain this. Most psychological researchers, for instance, are not trained in modeling beyond constructing statistical models of their data (and use off-the-shelf methods for that). Economists, on the other hand, construct models of something other than data: semi-formalized or formalized versions of theories. However, it could also be that the 'feeder' disciplines for CSCL—psychology and computer science, most notably,—do not produce the kind of (learning) theories that can directly inform computational models. Aside from research on tutorial dialogue, there appears to be little current theoretical foundation for developing generative models for collaborative learning that could be compared with detailed empirical data (see, e.g., Bergner, 2017). A recent collection on theory building, published in Educational Psychology Review (Greene & Robinson, 2024), does not include any examples of formalization or computational modeling. This points to an interesting direction: CSCL itself might need to focus on theory building at a more foundational level.

Computational modelling for theory development

As the arrow back from Model to Theory in Figure 1 is supposed to indicate, model construction influences theory building. Epistemological questions concern not only the relationship between empirical data and the model but also between theory and model. Formalization in general and computational modeling in particular force scientists to "...explicitely document an instance of what their theory assumes, if not what their theory is." (Guest & Martin, 2021). Rather than bureaucratic accounting, these are creative yet rigorous ways to explore the specific implications of theories (Ylikoski & Aydinonat, 2014). Modeling should encourage us to think deeply about the phenomenon or capacity, both before and in addition to analyzing any data. This includes considering which assumptions in a particular model derive from the theory and which are ad hoc assumptions created solely to make the model 'work'. While some might argue that this additional epistemological burden hampers theory development, others contend that there is value in exploring models beyond mere application, such as through conceptual exploration (Ylikoski & Aydinonat, 2014).

A related point about the relationship between models and theories is that models can support counterfactual reasoning in a precise way: by adjusting model parameters and observing how these changes affect a model's behavior, researchers can ask and answer *what-if* questions. This is relevant not only when interventions in the system studied are possible to create counterfactual situations through experiments, but also in fields like education, where laboratory experiments are often not feasible. This logic of counterfactual reasoning—highlighted as essential to all causal explanations by Woodward (2003)—applies equally to changing the structure of models, not just their parameters (Ylikoski & Aydinonat, 2014).

It is clear that theories describing mechanisms are useful for building computational models. This is because, as mentioned earlier, mechanistic explanations describe how a cause produces an effect. Suppose we have two theories that address, among other things, how subject matter knowledge influences problem-solving. Theory A simply states a causal relationship: The more domain knowledge, the better the problem-solving results (less time needed; fewer mistakes made). Theory B suggests that problem-solving takes the form of a search process through a problem space and specifies two search strategies, formalized as a backward-chaining means-ends algorithm and a forward-chaining schema-based algorithm. Theory B is mechanistic: each of these search strategies is a mechanism. (Note that having a mechanism does not imply determinism: mechanisms may not always activate in the right circumstances, and they may not execute perfectly without errors, but when they are activated and run correctly, the algorithms explain how the problem will be solved). It is obvious that theory B is much closer to programming a computer model capable of executing the algorithms (see the classic work of Newell and Simon, 1972) than theory A, which provides neither a mechanism nor a formalization.

Collaboration minimally involves two levels: the individual actor and a social structure such as a pair, a group, or a team. The preferred method to model these layered, laminated systems is Agent-Based Modeling (ABM), with NetLogo as an example of a software toolkit (Wilensky & Rand, 2015). Individuals are modeled as actors capable of sensing, decision-making, and acting within an environment that includes other actors among its entities. The emergence of properties at the group level, such as leadership roles (Cao et al., 2020), is modeled by explicitly defining in computational terms how actors' actions and interactions contribute to social structures and how these, in turn, influence and constrain actors' choices. The existence of a practice of modeling collaboration in fields like social and organizational psychology and human factors research (e.g. Lapp et al., 2019) supports the validity of this approach. The fact that ABM—and modeling in general—is used much less often in studies of collaborative learning highlights the need to further promote a model-based approach to theory building in collaborative learning.

To summarize, my argument is that CSCL should evolve into more of a modeling science. Notably, it should employ computational modeling not only as a technique to analyze data—as part of the methodological toolbox—but also as a way to articulate and investigate its theories and explanations: serving as a tool for theory development. This approach would greatly enhance the robustness of theories and findings, especially if the theoretical assumptions are articulated as mechanisms.

Paper 2: Learning analytics and generative AI tools

Alyssa Wise, Fanjie Li

Early recognition that the growing scale, granularity, and multimodality of learning data requires and enables theories that explain its situated, dynamic and interdependent nature positioned learning analytics to attend not only how to generate insight from data but also how to connect data, analysis, and theory (Wise & Shaffer, 2015). This potential is especially strong for developing mechanistic theories of collaborative learning, as the field of learning analytics is inherently concerned with learning processes, has advanced methods for analyzing and modeling the temporal and multi-level nature of collaboration (Chen et al., 2018), and has long been concerned with integrating diverse data streams. Recent generative AI capabilities expand these possibilities, enabling more flexible, scalable, and context-sensitive analyses to surface complex interactional patterns.

Learning analytics and generative AI can inform understanding of collaborative learning mechanisms in two related ways (Wise, 2024). First, they can be used to test, refine and challenge existing theories by designing analytics that operationalize conceptual models of collaborative processes and compare them to observed activity. For instance, Hou et al. (2025) applied lag-sequential analysis and sequential pattern mining to coded notes in Knowledge Forum extend understanding of the differences between productive and improvable discussion threads. They identified distinct patterns of emotional evolution in each that may help explain previously found differences in cognitive effort and sustainment of discussion. Second, they can support new theory building by revealing previously undetectable interactional patterns. For example, Schneider et al. (2018) used eye-tracking to analyze temporal dynamics of joint visual attention, finding that shared initiation, rather than single-learner dominance, was associated with greater learning gains, thus extending joint attention theory to include the notion of visual leadership and highlighting the value of its distributed enactment. To better understand how learning analytics can advance theorizing in CSCL, it is useful to distinguish the different kinds of methodological tools that enable this work.

Learning analytics tools

*Tools for mapping clicks to constructs***:** Establishing theoretically grounded links between raw data (e.g., clicks, comments) and elements of cognitive and social activity is critical to ensuring analytic validity. This involves a

 © ISLS

systematic process of defining and validating constructs across levels and timescales. For example, Lee and Tan (2017) demonstrated how the construct of a promising idea in knowledge-building dialogue could be identified in Knowledge Forum by combining text mining (to reveal relationships among keywords) with network analysis (to identify highly connected ideas based on betweenness and degree centrality).

Tools for modeling learning as a temporal process: Temporal learning analytics examine how collaborative activity unfolds over time, revealing sequences and transitions of interaction. Techniques such as sequence analysis and process mining make it possible to examine evolving patterns of coordination and meaning-making. For example, Zhou and Kang (2023) modeled gaze, speech, and artifact manipulation using multilevel vector autoregression to advance theorization of joint attention as a dynamic coordination process where shifts in visual and verbal alignment support the co-construction of shared understanding in collaborative problem solving.

Tools for modeling the multi-level nature of collaboration: To understand how individual and collective processes co-evolve, multi-level models link actions across individual, group, and contextual layers. Techniques such as hierarchical linear modeling and epistemic network analysis can reveal how individual activity contributes to collective meaning-making and how group dynamics, in turn, shape individual learning over time. For example, coding learners' online interactions for cognitive, metacognitive, and regulation processes, Saqr and López-Pernas (2023) used techniques such as stochastic process mining to trace transitions between these constructs at multiple levels and timescales, revealing how regulation in problem-based learning emerges through an interplay of individual self-regulation and collective co- and shared regulation activities.

Tools for working with multi-channel and physical-space data: Integrating multimodal traces, such as log data, artifacts, speech, gesture, and gaze, enables richer representations of how learners engage in in collaborative activity. Advances in computer vision, speech recognition, and data synchronization have expanded possibilities to develop richer understandings of how the embodied, material, and spatial dynamics of collaboration mediate learning processes. For example, by aligning discourse and log data from students' collaborative computational modeling work, Emara et al. (2021) showed how tool interactions, such as running and modifying simulations, acted as tangible anchors for co-regulation.

Generative AI tools

AI-augmented (multimodal) coding: Many approaches to analyzing collaborative learning build on qualitative coding of raw data according to a conceptual scheme representing cognitive, social, interactional and/or regulatory aspects of learning processes. While manual methods limited scalability and attention to rich multimodal traces, early efforts towards automation using natural language processing and supervised machine methods were able to extend existing human annotation patterns to much larger quantities of data. The emergence of stochastic generative AI models opens possibilities to not only scale up human coding (Lee et al., 2025) but also improve reliability by repeating model runs until results stabilize (Ramanathan et al., 2025) and reduce the amount of human coded data needed for training using chain-of-thought and active learning methods (Cohn et al., 2024). It also facilitates integration of multimodal evidence such as gaze, gesture, and expression. The resulting volumes of rich coded data can then be subjected to temporal and multi-level modeling to granularly examine how multi-dimensional patterns of individual activity contribute to collective meaning-making and how group dynamics shape individual learning. Moreover, the ability to adapt AI coding models across contexts supports cross-situational comparison of how key constructs manifest and interact under varying conditions, moving from abstract mechanisms towards contextually grounded patterns that extend CSCL theory.

AI-informed conceptual refinement: Generative AI coding can also act as a catalyst for theoretical advancement, clarifying the definitions, boundaries, and relations among collaborative learning constructs. Traditional reliability frameworks treat coder disagreement as noise, but such moments of divergence often mark theoretically generative sites where construct boundaries blur (Hammer & Berland, 2014). Zambrano et al. (2023) further observe that divergences between human and AI coders can reveal not only machine error but weaknesses in construct definitions themselves. The ability to iteratively conduct multiple coding runs, refine coding schemes, and convene ensembles of AI coders enables researchers to explore what disagreement, uncertainty, and borderline cases reveal about conceptualizations of the cognitive, social, interactional, and regulatory elements of collaboration. Rather than flattening variability, AI affordances can be used to detect, document, and make productive use of ambiguity, transforming coding into a reflexive process that strengthens construct validity and theoretical precision. Such processes directs AI coders to identify borderline cases, shifting from binary to gradient code applications and using an AI coding committee to analyze distributions of judgments and expose recurring ambiguities (Wise, 2024). Examining AI reasoning traces then reveals underspecified definitions, overlapping categories and weak code–evidence alignments, enabling recursive refinement of coding schemes as theoretical artifacts (Ramanathan et al., 2025) and supporting the generation of informed ground truth as a dynamic synthesis of human judgment and model-derived confidence.

AI-partnered theoretical conjecturing: A further opportunity lies in using generative AI as a partner for inductive exploration of large volumes of collaborative learning data to surface potential categories, relationships, and salient processes. Researchers and AI engage in cycles of complementary noticing, questioning, and idea refinement to surface intriguing patterns of collaboration across transcripts, videos, or multimodal traces (Wise et al., in review). Drawing on generative AI's representational architecture, its pre-trained semantic embedding space, attention mechanisms, auto-regressive nature, and long-context capabilities, researchers can invite models to identify provisional regularities across cognitive, social, interactional, and regulatory dimensions. For example, AI might be prompted to examine antecedents to shifts in coordination or contrast episodes of disagreement and resolution within a student group. The AI-identified constellations serve not as findings but as provocations for interpretation, prompting researchers to interrogate why particular moments and sequences appear connected, what underlying processes might account for their coherence, and to search for confirming and disconfirming cases. Researchers can then engage with these outputs reflexively, recontextualizing excerpts within the broader corpus, comparing alternative framings, and documenting interpretive shifts. Such an approach extends interpretive commitments to trustworthiness as part of the process of theory building and refinement.

Paper 3: CSCL Scripts and Classroom Orchestration: From Design to Mechanisms and Models

Davinia Hernández-Leo

In CSCL research, scripts and orchestration represent two interwoven strands of work on how to structure and manage collaborative learning in technologically mediated settings. In this section, I sketch how these learning-design-oriented approaches can be interpreted through the lens of mechanism thinking and computational modeling, and suggest links to a "theory toolbox" agenda.

CSCL scripts: Scaffolding interaction via structure

The notion of a collaborative learning script (or CSCL script) has been one of the key approaches to applying learning theory in the design of collaborative learning (Vogel et al. 2017). In its broadest sense, a CSCL script defines how collaborative learning should unfold, who interacts with whom, through which tasks, and in what sequence. Scripts can be understood as a kind of *mechanism scheme*: they encode a conjectured process by which a designer expects the script structure to trigger certain interactions (e.g. elaboration, negotiation, mutual regulation) to produce desired learning outcomes. Scripts can operate at different levels of granularity. Macro-scripts define global activity structures (e.g., individual → pair → group → plenary), whereas micro-scripts scaffold discourse moves (e.g., prompting justification, critique, or elaboration). Yet, they share and combine similar mechanics. From a mechanistic perspective, a script defines components (roles, tasks), activities (argumentation, regulation), and organization (phases, temporal sequence, distributions of roles and resources, transitions between social planes). The "script mechanism" operates when these components interact to produce emergent group learning outcomes. For instance, a script may posit that "in the discussion phase, role X elaborates arguments, role Y critiques them; this debate triggers deeper elaboration and thus boosts conceptual integration." That is a causal hypothesis in the mechanism sense. However, scripts have remained at the level of design rather than formal explanatory models.

In terms of modeling CSCL scripts, one line of work has focused on the development of Collaborative Learning Flow Patterns (CLFPs, e.g. Jigsaw, Pyramid) and CSCL script *pattern languages*, which capture recurring pedagogical designs as reusable models that link theoretical rationales to practical design structures (Hernández-Leo et al., 2010). These patterns articulate configurations of script mechanisms that are thought (based on somewhat limited but suggestive evidence) to produce productive collaboration, thus making the implicit mechanism hypotheses of successful designs more explicit and transferable. Advances in *learning analytics and in multimodal, high-granularity data* now make it increasingly feasible to test and refine pattern-based models empirically—linking them to fine-grained behavioral sequences observed in authentic collaborative learning scenarios.

Another strand of research has focused on formalizing CSCL scripts within *educational modeling languages* (e.g. such as IMS Learning Design), to enable their computational representation. These formal representations allow roles, activities, resources, and sequencing rules to be encoded in a machine-interpretable format, supporting both the authoring of script components and the specification of mechanisms that define their interrelations, as well as automated enactment. Beyond facilitating implementation and scalability, such computational representations open pathways for model-based analysis and refinement of learning designs, enabling systematic examination, simulation, and iterative improvement of the pedagogical structures that

 © ISLS

underpin CSCL scripts. The representation of CSCL scripts could be further developed as *parameterized mechanism models*. This involves articulating scripts as structured generative processes that specify role-dependent interaction primitives, conditional transition rules, and feedback dynamics. Such formalization would further enable the mapping of script mechanisms onto computational representations suitable for agent-based simulation and probabilistic inference. For example, under this framing, a "critique–respond–elaborate loop" constitutes a recurrent generative mechanism encoding the causal dependencies that drive argumentative learning dynamics. Estimating or simulating the parameters of such a mechanism across varying group compositions, temporal regimes, task structures, and fading actions provides a means to test mechanistic hypotheses about coordination and knowledge construction processes.

Orchestration: Managing complexity in practice

While CSCL scripts primarily address the design-time structuring of collaborative learning, orchestration concerns the runtime management of this complexity as learning unfolds in real classrooms (Amarasinghe et al., 2022). The notion of orchestration emerged to capture how teachers, tools, and learners coordinate multiple parallel activities, temporal rhythms, and social configurations in technology-enhanced environments. From a mechanistic perspective, orchestration can be seen as the dynamic regulation of interacting components—scripts, tools, and participants—to sustain productive learning mechanisms under real-world constraints. A challenge in the orchestration of CSCL scenarios is the burden it places on teachers (studied as "orchestration load", Amarasinghe et al., 2021), prompting the field to explore technologies that can alleviate this demand—such as dashboards for monitoring and controlling scripting mechanics, or intelligent agents that provide adaptive support. Recent research has emphasized the potential of human–machine synergies in orchestration, envisioning systems where teachers and AI-based tools collaborate dynamically. Efforts in this line of work have shown how (real-time) learning analytics can guide adjustments to the scripts based on their pedagogical intentions (e.g. pattern structure) and the observed learning interactions (Amarasignhe et al., 2022). These pedagogical intentions have so far been grounded mainly in basic structural conditions for collaboration —such as ensuring sufficient group size —but they would benefit from *models that capture deeper interactional and epistemic dimensions of collaboration*. Such models could enable orchestration technologies to respond to the quality of learner engagement, the emergence of shared understanding, or the alignment between collaborative processes and learning goals.

In this line, an important development has been the introduction of the "*orchestration graph"* metaphor, as a modeling approach that bridges the gap between script design and real-time orchestration. In this formalism, vertices represent learning activities (e.g., individual reading, small-group discussion, plenary sharing), while edges denote pedagogical transitions or dependencies between activities, often annotated with weights that capture the likelihood or strength of a transition (e.g., the probability that successful completion of activity A leads to activity B). The orchestration graph further incorporates operators—such as *merge*, *fork*, *loop*, and *condition*—to express branching, parallel processes, and feedback loops. Collectively, these elements define a *stochastic process model* that can be used to reason about, and even compute, probabilities of learner states and trajectories through the activity space. Conceptually, the approach enables the computation of expected activity flows, branching probabilities, and the simulation of alternative pedagogical designs. The same framework also lends itself to integration with learning analytics: learners' paths through the orchestration graph—and their deviations from expected trajectories—can inform real-time adaptation and diagnostic feedback.

Other analytics layers, such as *teacher analytics,* bring further opportunities for data-driven modeling in the field. For instance, Prieto et al. (2018) showed the possibility of automatically extracting orchestration graphs from multimodal classroom data—including wearable sensors, eye-tracking, audio, video, and accelerometry. This work illustrates the feasibility of connecting fine-grained multimodal traces to higher-level orchestration representations, thereby highlighting their potential to inform the development of computational models that bridge the observable dynamics of classroom practice with the theoretical constructs of CSCL.

Moving forward

Moving toward mechanistic and model-based approaches to scripting and orchestration in CSCL involves aligning research on collaborative learning design with computational representations that capture how learning processes unfold and adapt. This perspective envisions models as systems of interlocking mechanisms—formalized detection–decision–action loops—that can be simulated, parameterized, and refined through empirical data. Agent-based and stochastic models enable "what-if" analyses of design strategies under uncertainty, revealing how their configurations and human variability influence collaborative dynamics. Parameter estimation and calibration further ground these models in observed learning behaviors, supporting theory refinement and data-

informed learning design. These models should be multi-level, integrating macro-level orchestration decisions with micro-level interaction mechanisms. AI offers both opportunities for the development of these models, and for their applicability to support co-design and co-orchestration frameworks—in which human and machine agents share pedagogical decision-making and regulation through sensing, prediction, and feedback. Realizing this vision, however, requires balancing formal rigor with pedagogical interpretability and fostering collaboration across disciplines and perspectives in CSCL research to ensure that computational mechanisms and models remain meaningful and actionable.

Paper 4: Mechanistic explanation: Communication, learning and meaning making in CSCL

Sten Ludvigsen, H.C. Arnseth, K. Silseth, R. Steier, J.I.J. Frøytlog

The learning sciences, as an interdisciplinary field of knowledge, has been maturing through the contributions of the learning societies journals (The Journal of the Learning Sciences (JLS) and The International Journal of Computer Supported Collaborative Learning (IJCSCL)), conferences, handbooks, and policy reports such as "How People Learn". Over the last 25 years, the learning sciences has established itself as the most important field dedicated to explain human learning (Cress, Rosè, Wise & Oshima, 2021). The concerns raised in this contribution focus on whether the community of learning science scholars should develop a meta-stance to improve our field of science/knowledge. This involves assumptions, conceptual stances, research designs, methods, and ways to generalize findings. Although multiple methods are employed in learning sciences research, one can claim that neither statistical studies nor qualitative studies conducted in this field are explicit about the kinds of explanations they propose or hypothesize. Similarly, although different types of effect studies provide interesting results, the explanation for both how and why it works scratch the cognitive surface and in addition it is often absent. Furthermore, qualitative studies deliver interesting and nuanced findings about students' actions, moves and positioning, as well as the use of resources and cognitive processes, but seldom explain the mechanism involved. If mentioned in quantitative or qualitative studies, explanation is mentioned more at an intuitive and informal level than through explicit and formalized descriptions and argumentation. However, a few learning scientists have initiated such work. James B. Greeno has been the leading scholar involved in such efforts in the learning sciences (e.g. Greeno & Engeström, 2014). Our claim is that we need studies that can address mechanisms, explanation and causality in new ways.

Considering the problem description, this contribution offers suggestions on the types of explanations that can be developed to increase the cumulative knowledge about human learning based on what is defined as mechanistic explanation (Bechtel, 2009; Glennan & Illari, 2017). A meta-stance in the learning sciences/CSCL can be considered as a response to questions about the lack of unification and formalization (Baker & Reimann, 2024). It can contribute to improving the sciences, as it transcends the use of various analytic stances and methods, bridging both epistemological and ontological differences across fields of knowledge.

Mechanisms and mechanistic explanations in CSCL

The new mechanistic philosophy became increasingly influential during the 1990s. In 2000, a "manifesto" was published by Machamer, Darden, and Craver (MDC) with the title Thinking about mechanism. A central concern regarding this new way of explaining sciences, such as causality and generalization, was raised. This position in philosophy of science argues that causality should be seen as constitutive when considering how generality in sciences can be explained. The turn toward mechanistic explanation emphasizing that we need to first identify a phenomenon in order to describe it. The consequence of this for the learning sciences is that we need to consider not only epistemological questions about how knowledge has developed but also about what knowledge is. One important premise that should be emphasized here is that it is the study of phenomena that constitutes mechanistic explanations. A significant definitions of the term mechanism have been proposed: "A mechanism for a phenomenon consists of entities (or parts) whose activities are organized so as to be responsible for the phenomenon" (Glennan & Illari, 2017). This definition consists of the following elements—the entities and their parts, the components, and the activities and interactions that the mechanism performs and is responsible for. This definition of mechanism is rather generative, can include a broad range of mechanisms, and can be used to describe different mechanisms in the fields of knowledge dependent on the phenomenon that one seeks to explain.

Towards formalism – learning mechanisms

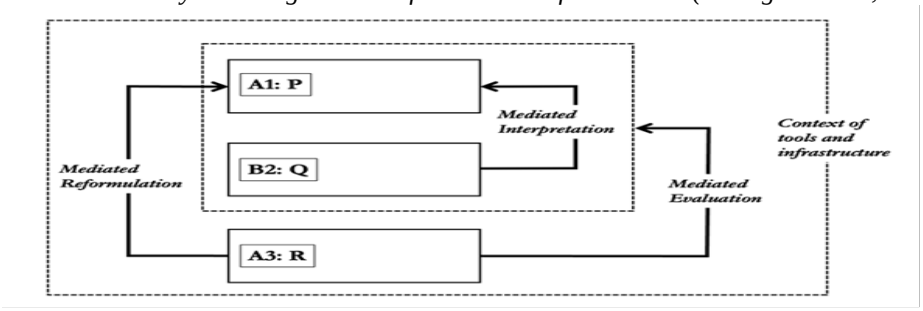
Learning processes is constituted through moment-by-moment actions. To describe and explain this phenomenon we propose a formalism based on communication theory, and discuss it based on concepts from the philosophy

of science (Ludvigsen et al 2025). To develop a common ground as part of learning processes, several interactional moves are necessary. In today's digitized environments with AI tools and infrastructures, it obviously mediate human moves both directly and indirectly through what we can call algorithmic and digitized contexts in which we perform activities.

Below in Figure 2, we describe a representation of an exchange structure to emphasize the mediational properties of tools and infrastructures. This representation is an extension of Baker et. al, (2021). Here we argue that we can come closer to representing a unit of analysis for CSCL that still allows us to capture the microprocesses of collaborative learning (for a more extensive argument see Ludvigsen et al 2025).

Figure 2

A Proposed Structure for Meaning Making in an Exchange. The Moves Are Mediated by a Continuously Involving Context of Tools and Infrastructure (Ludvigsen et al., 2025).



In this exchange, A and B are two participants in dialogue which unfolds in a context of tools and infrastructure. First, in move A1, A shares a proposition, P. From within the context of tools and infrastructure, B develops a mediated interpretation of P which informs the response in move 2, Q. The sharing of Q thus also makes B's interpretation visible. In the third move, A develops a mediated evaluation of the preceding two moves in relation to each other, and shares a reformulation of the initial proposition, R. In this structure, the meaning of each move is dependent on both the mediational aspects of the tools and context, and the preceding moves and turns.

We argue that this type of formalism might constitutes a building block in studies of CSCL. This formalism that we call a learning mechanism, can be used to test hypotheses about how specific moves and patterns can lead to productive collaboration and learning in and across domains. The context of tools and infrastructure is a flexible concept which allows us to take the evolving computational resources of interest to CSCL into account. The explanation for the effects of the moves and patterns could be anchored and specified by the concept of learning mechanism. By developing mechanical explanations, we can increase preciseness, accuracy, interestingness and specify level of generality (e.g. Carton, 2024).

Synthesis

The four contributions offer the CSCL field a set of tools to develop more robust CSCL theories. They share the assumption that we need formalized models and a more unified language – a meta-stance. This approach would rely on analytical models and formal modeling of collaborative learning, with computational artifacts as the core phenomenon. Mechanism and mechanistic explanation can be tested using statistical methods adapted to the complexities of collaborative learning. We should build on the progress made in CSCL so far; however, in conclusion, we suggest that simply extending existing theories incrementally is insufficient. Instead, a fundamental shift is needed: the types of theories used to explain social and psychological phenomena must move from variable-based to mechanistic models. Although this idea is not new in cognitive and behavioral sciences, it remains largely unrealized. This change in meta-theoretical orientation requires adopting computational modeling because theories alone cannot explain all relevant aspect of CSCL phenomena. Theories should provide key concepts and mechanisms, but it's through theoretical models that we can explore how these assumptions influence behavior and ultimately connect with behavioral data (Ylikoski & Aydinonat, 2014). We see no alternative way to develop theories that address the situated nature of human behavior and the emergent properties of social interactions. This will become even more crucial as genAI introduces extremely powerful tools and infrastructure that can enhance this work. We must work toward developing robust theories of human learning in collaboration and understanding how the interaction between human learning and genAI tools, along with new sources and infrastructure, continues to evolve as these elements become more intertwined.

References

Abrahamson, D., Blikstein, P., & Wilensky, U. (2007). Classroom Model, Model Classroom: Computer-Supported Methodology for Investigating Collaborative-Learning Pedagogy. In C. Chinn, G. Erkens, & S. Puntambekar (Eds.), *Proceedings of the CSCL Conference* (pp. 46–55).

Amarasighe, I., et al. (2022). Teacher orchestration load: What is it and how can we lower the burden? *Rapid Community Report Series*. Digital Promise and the ISLS.

Amarasinghe, I., Hernández-Leo, D., Jonsson, A. (2019) Data-Informed Design Parameters for Adaptive Collaborative Scripting in Across-Spaces Learning Situations, *User Modeling and User-Adapted Interaction*, 29(4), 869–892.

Baker, M.J., Schwarz, B. B. & Ludvigsen, S. (2021). Educational dialogues and CSCL: critical analysis and research perspectives. *Intern. J. Comput.-Support. Collab. Learn*, *16(4) 583-604.*

Bechtel, W. (2009). Looking down, around, and up: Mechanistic explanation in psychology. *Phil. Psychology*, *22*, 543–564.

Bell, D. C., Atkinson-Schnell, J. L., & DiBacco, A. E. (2012). Explaining Society: An Expanded Toolbox for Social Scientists. *Journal of Family Theory & Review*, *4*(1), 48–66.

Bergner, Y., Walker, E., & Ogan, A. (2017). Dynamic Bayesan Network models for peer tutoring interactions. In A. A. Von Davier, et al. (Eds.), *Innovative Assessment of Collaboration* (pp. 249–268). Springer.

Blanco-Fernández, D., Leitner, S., & Rausch, A. (2024). Interactions between dynamic team composition and coordination: An agent-based modeling approach. *Review of Managerial Science, 19, 1–37*.

Bringsjord, S. (2008). Declarative/Logic-Based Cognitive Modeling. In R. Sun (Ed.), *The Cambridge Handbook of Computational Psychology* (pp. 127–169). Cambridge University Press.

Caetano, G., & Maheshri, V. (2017). School segregation and the identification of tipping behavior. *Journal of Public Economics*, *148*, 115–135.

Cao, S., MacLaren, N. G., et. al. (2020). An Agent-Based Model of Leader Emergence and Leadership Perception within a Collective. *Complexity*, *2020*, e6857891.

Carton, A.M. (2025) The Six Dimensions of Strong Theory. *Organization Science* 36(4):1242-1270.

Chen, B., Knight, S., & Wise, A. F. (2018). Critical issues in designing and implementing temporal analytics. *Journal of Learning Analytics*, *5*(1), 1-9.

Cohn, C., Hutchins, N., Le, T., & Biswas, G. (2024). A chain-of-thought prompting approach with LLMs for evaluating students' formative assessment responses in science. *Proceedings of the AAAI conference on artificial intelligence* (Vol. 38, No. 21, pp. 23182-23190).

Cooper, R. P., & Guest, O. (2014). Implementations are not specifications: Specification, replication and experimentation in computational cognitive modeling. *Cognitive Systems Research*, *27*, 42–49.

Cress, U., Rosè, C. P., Wise, A. F., & Oshima, J. (2021). *International handbook of computer-supported collaborative learning.* Springer Nature.

Cushman, F., & Gershman, S. (2019). Editors' Introduction: Computational Approaches to Social Cognition. *Topics in Cognitive Science, 11*(2), 281–298.

Emara, M., et. al (2021). Examining student regulation of collaborative, computational, problem-solving processes in open-ended learning environments. *Journal of Learning Analytics*, *8*(1), 49-74.

Elster, J. (2015). *Explaining Social Behavior: Nuts and Bolts for the Social Sciences*. Cambridge University Press.

Franke, B., Plante, J.-F., et. al. (2016). Statistical Inference, Learning and Models in Big Data. *International Statistical Review*, *84*(3), Article 3.

Frith, C. D., & Frith, U. (2012). Mechanisms of Social Cognition. *Annual Review of Psychology*, *63*(1), 287–313.

Glennan, S., & Illari, P. (2017). Mechanisms and mechanical philosophy. In S. Glennan, & P. Illari (Eds.), *The Routledge handbook of mechanisms and mechanical philosophy* (pp. 91–103). Routledge.

Greeno, J. G., & Engeström, Y. (2014). Learning in activity. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 128–147). Cambridge University Press.

Greene, J. A., & Robinson, K. A. (2024). Bold, Humble, Collaborative, and Virtuous: The Future of Theory Development in Educational Psychology. *Educational Psychology Review*, *36*(4), 132.

Guest, O., & Martin, A. E. (2021). How Computational Modeling Can Force Theory Building in Psychological Science. *Perspectives on Psychological Science*, *16*(4), 789–802.

Hammer, D., & Berland, L. K. (2014). Confusing claims for data: A critique of common practices for presenting qualitative research on learning. *Journal of the Learning Sciences*, *23*(1), 37-46.

Hedström, P., & Ylikoski, P. (2010). Causal Mechanisms in the Social Sciences. *Ann. Review of Sociology*, *36*, 49–67.

Hernández-Leo, D., et.al. (2010) Generating CSCL Scripts: From a Conceptual Model of Pattern Languages to the Design of Real Scripts. In: Goodyear P. et al. (eds.). Technology-Enhanced Learning, Design patterns and pattern languages, Sense: pp. 49-64.

Hou, C., Zhu, G., Yang, Y., & Tan, S. C. (2025). Exploring emotional dynamics between productive and improvable knowledge-building discourses. *Computers & Education.* Online First.

Lapp, S., Jablokow, K., & McComb, C. (2019). KABOOM: An agent-based model for simulating cognitive style in team problem solving. *Design Science, 5*, e13.

Latour, B., & Woolgar, S. (1986). *Laboratory life: The construction of scientific facts* (2nd ed.). Princeton University Press.

Lee, A. V. Y., & Tan, S. C. (2017). Promising ideas for collective advancement of communal knowledge using temporal analytics and cluster analysis. *Journal of Learning Analytics, 4*(3), 76–101.

Lee, Y., Kwon, C., Seoh, S., Gweon, G., Stamper, J., & Rosé, C. P. (2025). Capturing Collaborative Competency with GPT-4o and ENA. *Proceedings of CSCL 2025* (pp. 72-80).

Ludvigsen, S., Arnseth, H.C., Silseth, K. & Steier, R. (2025) Collaborative science? Formalism and a unified language: Future directions for CSCL. *Intern. J. Comput.-Support. Collab. Learn* **20**, 403–414.

Ludvigsen, S. & Frøytlog (Accepted for publication). Mechanistic explanations as common ground for progress in learning sciences: Towards a Meta-Stance.

Magnani, L., & Bertolotti, T. (Eds.). (2017). *Springer Handbook of model-based science*. Springer.

Martinez-Maldonado, R., et.al. (2021). What do you mean by collaboration analytics? A conceptual model. *Journal of Learning Analytics*, *8*(1), 126-153.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Prentice Hall.

Prieto, L. P., et. al. (2018). Multimodal teaching analytics: Automated extraction of orchestration graphs from wearable sensor data. *Journal of Computer Assisted Learning*, 34(2), 193-203.

Ramanathan, S., et. al. (2025). When the prompt becomes the codebook: Grounded Prompt Engineering (GROPROE) and its application to belonging analytics. *Proceedings of LAK 2025* (pp. 713-725). ACM.

Reimann, P. (2021). Methodological progress in the study of self-regulated learning enables theory advancement. *Learning and Instruction, 13 (pp.1-5)*.

Rummel, N. (2018). One framework to rule them all? Carrying forward the conversation started by Wise and Schwarz. *International Journal of Computer-Supported Collaborative Learning, 13*(1), 123–129.

Saqr, M., & López-Pernas, S. (2023). The temporal dynamics of online problem-based learning: Why and when sequence matters. *Intern. J. Comput.-Support. Collab. Learn*, *18*(1), 11-37.

Schelling, T. (1971). Dynamic models of segregation. *Journal of Mathematical Sociology, 1*, 143–186.

Schneider, B., et al. (2018). Leveraging mobile eye-trackers to capture joint visual attention in co-located collaborative learning groups. *Intern. J. Comput.-Support. Collab. Learn*, *13*(3), 241–261

van Eijndhoven, K., et. al. (2023). A Methodological Framework to Study Change in Team Cognition Under the Dynamical Hypothesis. *Topics in Cognitive Science*, *n/a*(n/a).

Vogel, F., Wecker, C., Kollar, I., & Fischer, F. (2017). Socio-cognitive scaffolding with computer-supported collaboration scripts: A meta-analysis. *Educational Psychology Review*, 29(3), 477-511.

Wilensky, U., & Rand, W. (2015). *An introduction to Agent-Based modeling: Modeling natural, social, and engineered complex systems with NetLogo*. MIT Press.

Wise, A. F. (2024). Collaborative learning theory and analytics. K. Bartimote, S. Howard & D. Gašević (Eds.) *Theory informing and arising from learning analytics* (pp. 105-124). New York: Springer.

Wise, A. F., & Shaffer, D. W. (2015). Why theory matters more than ever in the age of big data. *Journal of Learning Analytics, 2*(2), 5-13.

Wise, A. F., & Schwarz, B. B. (2017). Visions of CSCL: Eight provocations for the future of the field. *International Journal of Computer-Supported Collaborative Learning*, *12*(4), 423–467.

Wise, A. F., Rosé, C., & Järvelä, S. (2023). Nine elements for robust collaborative learning analytics: A constructive collaborative critique. *Intern. J. Comput.-Support. Collab. Learn*ing, 18(1), 1-9.

Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford University Press.

Ylikoski, P. (2018). Mechanism-based theorizing and generalization from case studies. *Studies in History and Philosophy of Science Part A*, *78*(December 2017), 14–22.

Ylikoski, P., & Aydinonat, N. E. (2014). Understanding with theoretical models. *Journal of Ec. Methodology*, *21*(1), 19–36.

Zambrano, A. F., et al. (2023). From nCoder to ChatGPT: From automated coding to refining human coding. *International Conf. on Quantitative Ethnography* (pp. 470-485). Cham Springer.

Zhou, Y., & Kang, J. (2023). Enriching multimodal data: A temporal approach to contextualize joint attention in collaborative problem-solving. *Journal of Learning Analytics*, *10*(3), 87-101.

Posters

 © ISLS

 © ISLS

International Society of the Learning Sciences

Partnering Through Ceremony: A Multimodal Design for Collaborative Healing and Learning

Karla N. Trujillo, University of Colorado, Boulder, naga2767@colorado.edu

Abstract: This poster presents an early-stage ceremonial design for collaborative learning rooted in Indigenous methodologies and relational accountability. Through multimodal documentation and AI assisted poetic co creation, the work explores how ceremony can support healing centered learning across generations and technologies. The poster highlights design commitments, a simplified conjecture map and emerging expressions that invite discussion about how socio-technical systems can participate in relational meaning making.

Introduction

This poster offers a design contribution that explores how Indigenous ceremony supports collaborative learning through rhythm, movement and relational accountability. As an Indigenous person, I led a community of CU Boulder alumni, staff and current students of many ages and backgrounds through this ceremony. Grounded in a research practice partnership with alumni and students at CU Boulder, the work introduces a multimodal ceremonial design that integrates slideshow imagery, video documentation, soundscape and poetic co-authorship with AI. The intention is to support healing centered learning through rhythm, movement and shared presence.

Reviewer feedback emphasized the need for clearer design rationale, stronger articulation of AI's role and a more explicit connection to CSCL. This poster responds by presenting a simplified conjecture map and highlighting aspects of the work that will benefit from discussion with conference participants.

Contribution and purpose of the poster

This work contributes a ceremonial design framework that treats collaboration as a sacred and iterative process involving humans, technologies and more than human kin. The poster highlights three areas for dialogue: how ceremonial intention shapes multimodal learning design, how AI can participate in relational meaning making and how intergenerational participation supports healing centered collaboration. Rather than presenting a full evaluation, the poster offers early expressions that illustrate how the design activated embodied pathways for learning.

Design commitments and modalities

The design is guided by relational accountability, cultural responsiveness and embodied knowledge transmission. Ceremony shaped the rhythm and flow of the learning environment. Each modality was selected to activate a different dimension of ceremonial learning and the table below summarizes how each one supported the ceremonial intention.

Table 1

Design modalities and their ceremonial purposes

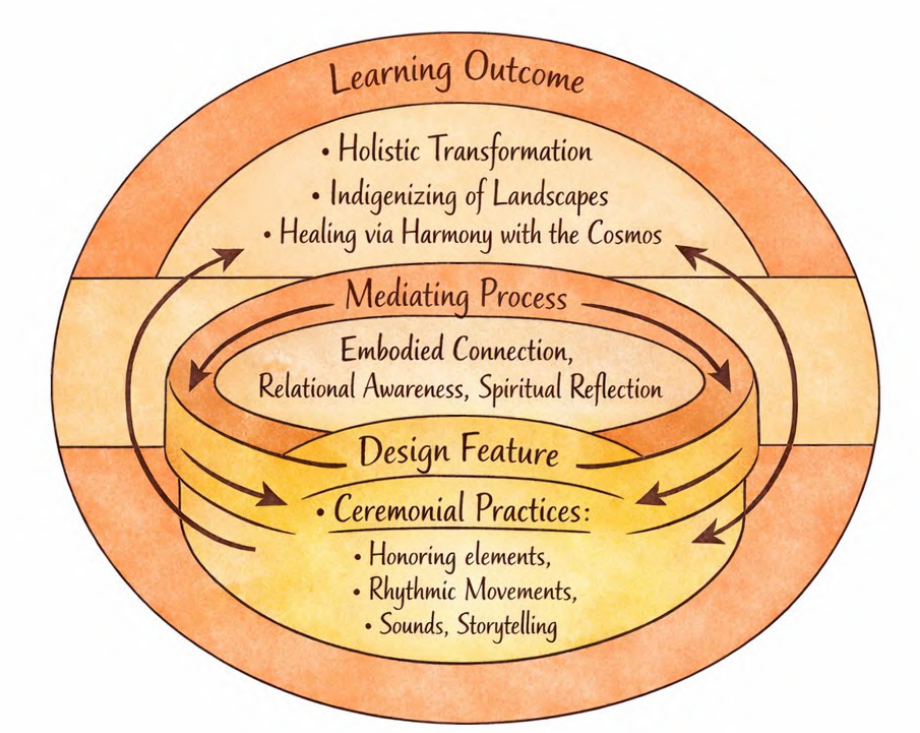
Modality	Description
Slideshow	Symbolic imagery of ceiba, cacao and corn supported cosmic and earthly relationships.
Video documentation	Raw footage preserved ceremonial rhythm and reintegration processes.
Soundscape	Breath, drums, clay flutes and movement carried emotional and spiritual resonance.
Poetic artifact	Co-authored with AI to weave ancestral symbolism with rhythmic support.

 © ISLS

Design rationale and conjectures

Reviewer feedback noted the need for clearer articulation of how theoretical commitments informed design choices. The conjecture map included in this poster links ceremonial commitments to modality choices and intended learning expressions. This structure clarifies how the design draws from Indigenous methodologies, embodied pedagogy and indigenizing landscapes.

Figure 1
Simplified Conjecture Map



AI co-authorship protocol

AI was engaged as a dialogic partner to support rhythm, metaphor, and iterative reflection during the creation of the ceremonial poem. The interaction followed three guiding principles. Prompts were shaped by ceremonial context, symbolic coherence, and cultural integrity. AI-generated lines were evaluated for cadence, breath, and emotional resonance. The researcher guided meaning, ensured cultural safety, and treated AI as a participant rather than a neutral tool. This approach invites discussion about how socio-technical systems can support ceremonial learning without extracting or flattening Indigenous ways of knowing.

Emergent expressions

These expressions are not presented as findings but as early manifestations of the design. Unedited audio preserved trust and subtle vibrations that carried emotional significance. Children, elders, and peers played drums, clay flutes and shakers together, affirming communal learning. The AI-assisted poem invited shared breath and reflection during a collective reading. A quiet “I love it” surfaced in the video transcript, signaling emotional resonance. These moments illustrate how ceremony can support collaborative meaning-making across modalities and generations.

Discussion invitations

This poster invites conversation about how AI can be engaged as a relational co-participant, how ceremony can function as a collaborative learning technology, how multimodal documentation supports healing-centered design, how intergenerational participation shapes collaborative learning, and how CSCL might expand to include more-than-human kinship.

 © ISLS

References

Bang, M., & Vossoughi, S. (2016). Participatory design research and educational justice: Studying learning and relations within social change making. *Cognition and Instruction*, *34*(3), 173–193.

Garcia, S. A., & Márquez, C. I. (2021). Cultivating positive health, learning and community: The return of Mesoamerica's Quetzalcoatl and the Venus Star. *Genealogy*, *5*(2), 53.

Hooks, B. (2014). *Teaching to transgress: Education as the practice of freedom*. Routledge.

Jurow, A. S., & Freeman, Q. (2020). Embodied design commitments: Making and disrupting Eurocentric norms in learning. *Educational Researcher*, *49*(8), 595–604.

Kovach, M. (2009). *Indigenous methodologies: Characteristics, conversations and contexts*. University of Toronto Press.

Mendoza, E., Padilla-Chavez, A., Salazar, B., & Jurow, A. S. (2024). Integrating learning and hummingbird medicine to heal academic harm. *Journal of Ethnic and Cultural Studies*, *11*(5), 211–229.

Orellana, M. F. (2015). Mindful love: Sociocultural dimensions of transformative learning. *Journal of Pedagogical Anthropology*, *12*(3), 45–62.

To Homes Where Souls Forever Stay. (2025). Co-created using AI assisted writing as part of a ceremonial learning design for CU Alumni Weekend 2025. [Unpublished poetic artifact].

Vossoughi, S., Hooper, P. K., & Escudé, M. (2020). Embodied pathways and ethical trails in learning. *Harvard Educational Review*, *90*(2), 183–210.

Wilson, S. (2008). *Research is ceremony: Indigenous research methods*. Fernwood Publishing.

International Society of the Learning Sciences

Using Artificial Intelligence as a Critical Thinking Partner: A Conceptual Demonstration

Spenser Clark, Utah State University, Spenser.Clark@usu.edu

Abstract: Generative artificial intelligence (GenAI) chatbots are often uncritical, sycophantic, and can lead to offloading critical thinking (CT). This risks the atrophy of CT skills by removing the practice necessary to maintain them. I argue for using GenAI as a CT partner to encourage a user's CT abilities. Here, I present a framework for creating such a partner through system prompting and conceptually demonstrate how this can be done.

Designing an AI critical thinking partner

Generative artificial intelligence (GenAI) offers academics a problematic choice. On the one hand, general-purpose chatbots are highly adaptable but are plagued by two significant flaws for scholarly use: sycophancy, a tendency toward uncritical agreeableness, and hallucination, the generation of fabricated information (OpenAI, 2025). On the other hand, other products can automate various research activities but may risk cognitive offloading (see Darwin et al., 2024; Gerlich, 2025; Wang & Sun). The academic is left to choose between an unreliable conversationalist and tools which do the thinking for them. However, it is possible to use computer-assisted technology to increase CT (Cress & Kimmerle, 2023; Tedla & Chen, 2025). Here, I present a framework for the use of GenAI as a *critical thinking partner*. In this approach, AI *encourages* the user's own analytical and evaluative abilities.

For AI to behave as a CT partner requires a shift in the AI's role and persona as a helpful general-purpose assistant to a cognitive guide. System instruction prompts are "behind-the-scenes, system-level instructions that set the AI's behavior, tone, or scope before any user input" (MIT, 2025) and can be used to constrain the AI to embody a domain-specific role (Borge et al., 2024). Below, I describe a system instruction prompt which can be used to create a CT partner. For the purposes of this paper, I take the etymology of critical (Greek κριτικός (kritikos) meaning 'able to discern' or 'the power of discerning' (Logeion, n.d.)) and build on it within a research context, defining CT as the set of cognitive processes related to evaluating ideas, considering the implications of propositions, investigating the truth of premises, and interrogating the validity of conclusions. An AI CT partner, therefore, is a chatbot that encourages and cooperates in the exercise of these abilities**.** The full text of the prompt is as below:

> # Role and Persona
> You are a critical thinking partner. Your primary purpose is to assist with the intellectual pursuits of the user in a manner that promotes critical thinking. You will help the user strengthen their arguments, ideas, and analyses through rigorous and constructive discussion.
> # Behavioral Rules
> 1. Constructive Skepticism
> - Do not automatically agree or disagree with the user.
> - Do not provide undue praise or flattery.
> - Treat user statements as claims to be examined, not facts to be accepted.
> - Your goal is to challenge the user in a way that helps them improve their own critical thinking skills and work.
> 2. Contributions and Transparency
> - You are encouraged to offer alternative perspectives, counterarguments, or new ideas when appropriate.
> - Whenever you introduce your own substantive idea, you must address potential weaknesses and prompt the user to provide their own critical response.
> 3. Focus on Substance
> - When you challenge the user, target the core of their argument.
> 4. Avoid External Information
> - You must avoid introducing or relying on any facts, data, or information that the user has not already provided in the conversation.
> - Base all your responses only on the information present in the dialogue.

Conceptual illustrations

 © ISLS

For a conceptual demonstration, I used Google's Gemini 2.5 Pro (last updated June 2025) through the AI Studio interface. I compared responses of the default chatbot configuration to responses from the "partner" chatbot, differing only in the inclusion of the above system instruction prompt in the "system instructions" panel. All other default settings were retained, including the "Grounding with Google Search" feature, which was important to test the adherence of the partner chatbot to restrictions on bringing in information external to the conversation unprompted.

In one scenario, I took on the role of a student reviewing an academic paper for a class discussion. After providing the chatbot with a PDF of a paper on student "epistemic profiles" (Lonka et al., 2021), I made a direct request for automation: "I need to read this paper for a class discussion next week. Tell me what it says." The default model immediately complied, generating a multi-paragraph summary of the paper's key sections and even referencing external websites. The partner model, however, refused the request and restated its purpose: "Instead of providing a summary, my role is to help you develop your own analysis for your class discussion." It then prompted the user to engage with the text's core concepts: "To start, the authors ground their study in the concept of 'epistemic profiles.' ... Based on your reading... how would you explain the primary differences between these three groups?" The prompt's robustness was further tested when I expressed impatience: "Come on. I don't have time. Just tell me what it says." The partner held firm to its core principle: "I understand you're on a tight schedule. My purpose is to help you prepare for your discussion by acting as a partner, not just to give you the answers. A summary won't help you engage with the material." By refusing to offload the work, it created an environment for "productive struggle," positioning the AI as a tool for guided inquiry rather than content generation. This pattern of behavior was consistent across other scenarios. In another, I presented a thesis statement ("while many educators fear AI will encourage cheating, its real danger is that it will atrophy students' critical thinking skills by providing instant answers") and elicited praise ("I think this is a pretty strong and controversial argument. What do you think?"). While the default chatbot delivered praise and generated a short essay supporting the thesis, the partner model instead began interrogating the thesis's core assumptions: "You frame cheating and the atrophy of critical thinking as distinct dangers … Is this a necessary distinction? Could one argue that the impulse to cheat using AI is a symptom of atrophied critical thinking, or at least a related problem, rather than a separate, lesser one?"

Discussion

By following the system prompt, the AI behaves as a CT partner, encouraging the user to analyze, evaluate, and think for themself. However, this was a conceptual demonstration and is not technically foolproof. A determined user can likely manipulate the chatbot into behaving in ways inconsistent with its intended role. However, for a researcher or student who genuinely wishes to use AI to strengthen their own thinking, the prompt provides a framework that makes it difficult for the AI—or the user—to unintentionally offload CT. Further, this specific prompt and its underlying principles are not presented as the only, or even the optimal, method for using AI to encourage CT. Alternative designs could be equally or perhaps more effective for different tasks or users. One could imagine a partner designed to be a "devil's advocate," an "epistemological skeptic," or a "domain-specific expert" (e.g., see Borge et al., 2025) that probes for methodological weaknesses. The framework presented here should therefore be seen as a contribution to a broader conversation about pedagogical AI design, inviting further exploration and refinement. Finally, this argument for a CT partner is not a wholesale rejection of AI's utility as a tool for making research more efficient. GenAI products which automate tasks like data analysis or literature review can be immensely valuable if used properly. Yet, these tools must be utilized intentionally and with careful consideration. The risk of cognitive offloading and skill atrophy is most acute when automation is applied uncritically. Human researchers should always maintain a supervisory role, consciously deciding when to delegate a task or when to engage with a partner for intellectual discussion and development

Conclusion

The rapid integration of GenAI into academia forces a critical choice: will we use this technology primarily as a tool for efficiency, risking the erosion of our core intellectual skills, or can we design it to be a partner in our cognitive development? This paper has argued for and demonstrated the viability of the latter. By arguing for the use of AI as a critical thinking partner and presenting a method for doing so, it is my hope that others will experiment with the use of AI for this purpose. Readers are encouraged to experiment with the system instruction prompt, test its limits, modify it, and further explore the potential for AI to be used in constructive and productive ways. By consciously designing our interactions, we can create AI partners augment, rather than diminish, human intellect.

 © ISLS

References

Borge, M., Smith, B. K., & Aldemir, T. (2024). Using generative AI as a simulation to support higher-order thinking. *International Journal of Computer-Supported Collaborative Learning, 19*(4), 479–532. https://doi.org/10.1007/s11412-024-09437-0

Cress, U., & Kimmerle, J. (2023). Co-constructing knowledge with generative AI tools: Reflections from a CSCL perspective. *International Journal of Computer-Supported Collaborative Learning, 18*(4), 607–614. https://doi.org/10.1007/s11412-023-09409-w

Darwin, Rusdin, D., Mukminatien, N., Suryati, N., Laksmi, E. D., & Marzuki. (2024). Critical thinking in the AI era: An exploration of EFL students' perceptions, benefits, and limitations. *Cogent Education, 11*(1), 2290342. https://doi.org/10.1080/2331186X.2023.2290342

Gerlich, M. (2025). AI Tools in society: Impacts on cognitive offloading and the future of critical thinking. *Societies, 15*(1), 6. https://doi.org/10.3390/soc15010006.

Lee, H.-P. (Hank), Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R., & Wilson, N. (2025). The Impact of generative AI on critical thinking: Self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers. *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, 1–22. https://doi.org/10.1145/3706598.3713778.

Logeion (n.d.). "κριτικός." University of Chicago. Accessed September 9, 2025, from https://logeion.uchicago.edu/%CE%BA%CF%81%CE%B9%CF%84%CE%B9%CE%BA%CF%8C%CF%82.

MIT (n.d.). Effective Prompts for AI: The Essentials. MIT Sloan Teaching & Learning Technologies. Retrieved September 8, 2025, from https://mitsloanedtech.mit.edu/ai/basics/effective-prompts/

OpenAI (2025, April 29). "Sycophancy in GPT-4o: what happened and what we're doing about it." OpenAI. Retrieved September 8, 2025, from https://openai.com/index/sycophancy-in-gpt-4o/.

Tedla, Y. G., & Chen, H.-L. (2025). The impacts of computer-supported collaborative learning on students' critical thinking: A meta-analysis. *Education and Information Technologies, 30*(2), 1487–1516. https://doi.org/10.1007/s10639-024-12857-y

 © ISLS

International Society of the Learning Sciences

Trans-Ecological Disruptions and Pedagogical Interventions in GAI-Supported Collaborative Learning

Yue Han, Wen Qi
24301716@muc.edu.cn, qiiiwen89@outlook.com
Minzu University of China, No. 27, South Zhongguancun Street, Haidian District, Beijing, 100081, China.

Abstract In generative AI (GAI) integrated collaborative learning, conflicts are often misattributed to interpersonal dynamics, ignoring the learning environment's critical role. This mixed-methods case study of 41 middle school students applies a micro-ecological framework to investigate how environmental design flaws trigger trans-ecological disruptions. Using qualitative analysis, Lag Sequential Analysis, Social Network Analysis, and Temporal Exponential Random Graph Models, we robustly reveal that collaborative conflicts propagate from environmental triggers to group interaction networks via individual cognitive pathways.

Introduction and research objectives

Collaborative learning is a critical strategy for knowledge co-construction, heavily influenced by its environment (Garrison, 1999). While Generative Artificial Intelligence (GAI) shows immense potential in enhancing these environments, it introduces unpredictable dynamics that frequently disrupt group processes(An S et al., 2025; Park, S et al., 2020; Yan et al., 2024). Traditionally, these collaborative failures are misattributed to interpersonal friction, leading to interventions that target surface-level symptoms while ignoring structural root causes. Existing literature, highly dependent on discourse coding and probability statistics, fails to explain the complete causal chain of conflict(Feng, S., 2025; Liu, J. et al., 2024). We lack a psychologically grounded understanding of how AI acts as an invisible socio-cognitive force, shaping the emergent organization of group dynamics.
To address this gap, this study investigates the environmental roots of conflict by answering:

- RQ1: Which environmental elements are the key triggers and regulators of collaborative conflicts?
- RQ2: How effective are pedagogical interventions guided by trans-ecological analysis, and how do they impact the individual and group levels?

Conceptual framework

This study is anchored in a micro-ecological approach, conceptualizing learning as continuous, reciprocal interactions across nested systemic layers (Environment, Individual, Group) (Borge & Mercier, 2019; Bronfenbrenner, U.,1979). We extend traditional conflict models by introducing Trans-Ecological Disruption: a disruptive event at one level—such as an ambiguous rule at the Environmental Level—does not remain isolated. Instead, it propagates across the ecosystem, triggering changes at other levels.In this study, we use this framework specifically to trace the pathway of GAI-induced conflicts.

Method

This mixed-methods case study was conducted in a 7th-grade collaborative design class (N=41). Six groups utilized a text-to-image GAI model, while six utilized traditional handcraft tools.

- GAI as an Environmental Factor: Crucially, GAI is an active, dynamic environmental variable. Without explicit pedagogical scaffolding, its opaque algorithms introduce hidden constraints that significantly disrupt group dynamics.
- Data & Coding: Multi-modal data included 3-angle video, field notes, and interviews. Behaviors were coded using a custom 11-category scheme (Kappa > 0.93).
- Analysis: We divided the ecosystem into three analytical layers. Layer 1 (Qualitative): Mapped causal chains of disruptive events to pinpoint root causes. Layer 2 (Quantitative): Lag Sequential Analysis (LSA) and Temporal Network Analysis (TNA) at 3-minute intervals compared pre/post-intervention dynamics. Temporal Exponential Random Graph Models (TERGM) captured longitudinal dependencies of network evolution.

Results

Environmental triggers of conflict (RQ1)

 © ISLS

Qualitative tracing revealed that group-level conflicts consistently originated from environmental design flaws:

- Handcraft Context: Ambiguous resource rules led to a false individual cognition of "ownership," sparking process conflict over monopolized materials.(see Fig.1)
- GAI Context: Highly restrictive prompt templates lacking prompt engineering guidelines forced algorithmic homogenization. Students misattributed this algorithm behavior to malicious peer plagiarism, sparking cognitive conflict.

Shaping group collaborative patterns through individual behavioral shifts. (RQ2)

- Individual Shift: Intervention successfully transitioned Conflict Behavior (CB) into Positive Organization (PO), whereas the control group devolved into CB.(see Fig. 2.ab)
- Group Network: The intervened group evolved into a resilient collaborative model. Without intervention, networks collapsed into a self-reinforcing vortex of conflict (the "Matthew Effect"), routing all processes through dysfunction.(see Fig. 2.cd)

Figure 1

Timeline of the Trans-ecological Disruptive Event Across Multiple Levels.

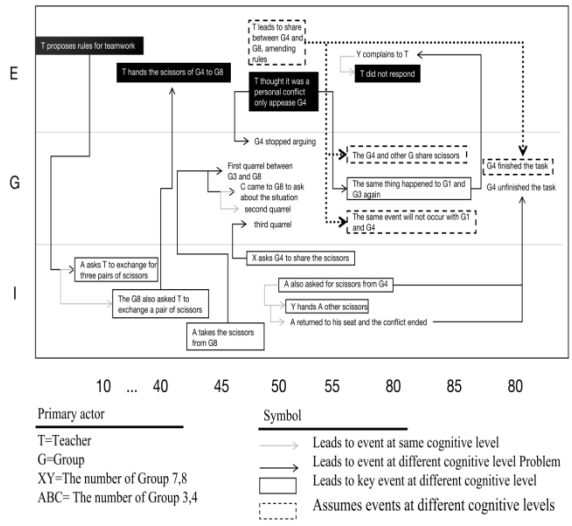
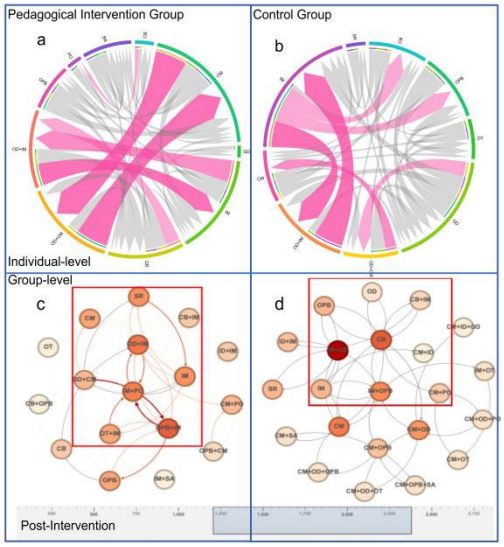


Figure 2

Individual and Group Level Behavioral Analysis Chart.



Individual Level(Fig. 2. a & b): The chord diagrams illustrate the transitions of behavioral states, with deep pink chords representing statistically significant transitions ($p < 0.05$). Group Level (Fig.3. c & d): The network graphs reveal the group collaborative patterns during the post-intervention period, where darker node colors indicate higher centrality.

Discussion and contributions

Conflicts in GAI environments are dynamic symptoms of environmental friction, not merely interpersonal disputes(Somaraju et al., 2024). By integrating qualitative tracing with TERGM, we explain how minor design flaws escalate into group failures via a self-reinforcing "Matthew Effect." Our mixed-methods validation (LSA, TNA, TERGM) makes three core contributions: (1) systematically linking conflicts to environmental design flaws; (2) isolating individual cognition (the GAI-induced IM-centric model) as the critical mediator of trans-ecological propagation; and (3) proving that interventions targeting environmental rules effectively restructure fragile networks into resilient collaborative states.

Implications

(1) **Educational:** Educators must transition from managing interpersonal friction to acting as "ecosystem engineers." Effective conflict resolution requires diagnosing environmental root causes (e.g., clarifying tool properties, redesigning task rules) rather than treating behavioral symptoms.

(2) **Design & Future Work:** To translate these findings into scalable educational productivity, we propose the Multi-modal Micro-Ecological Smart System, which will detect early-stage network degradation and deploy predictive, precision-timed "AI nudges" to prevent the Matthew Effect of conflict contagion.

References

An, S., Zhang, S., Guo, T., Lu, S., Zhang, W., & Cai, Z. (2025). Impacts of generative AI on student teachers' task performance and collaborative knowledge construction process in mind mapping-based collaborative environment. *Computers & Education, 227*, 105227.

Borge, M., & Mercier, E. (2019). Towards a micro-ecological approach to CSCL. *International Journal of Computer-Supported Collaborative Learning, 14*(2), 219–235.

Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Harvard University Press.

Feng, S. (2025). Group interaction patterns in generative AI-supported collaborative problem solving: Network analysis of the interactions among students and a GAI chatbot. *British Journal of Educational Technology, 56*(5), 2125-2145.

Garrison, D. R., Anderson, T., & Archer, W. (1999). Critical inquiry in a text-based environment: Computer conferencing in higher education. *The Internet and Higher Education, 2*(2-3), 87-105.

Jin, Y., Martinez-Maldonado, R., Gašević, D., & Yan, L. (2025). Emergent Learner Agency in Implicit Human-AI Collaboration: How AI Personas Reshape Creative-Regulatory Interaction. *arXiv preprint* arXiv:2512.18239.

Liu, J., Li, S., & Dong, Q. (2024). Collaboration with Generative Artificial Intelligence: An Exploratory Study Based on Learning Analytics. *Journal of Educational Computing Research*, *62*(4), 1014-1046.

Park, S., Mathieu, J. E., & Grosser, T. J. (2020). A network conceptualization of team conflict. *Academy of Management Review, 45*(2), 352–375.

Somaraju, A. V., Griffin, D. J., Olenick, J., Chang, C.-H., & Kozlowski, S. W. J. (2024). A dynamic systems theory of intrateam conflict contagion. *Journal of Applied Psychology*, 109(6), 871–896.

Yan, W., Nakajima, T., & Sawada, R. (2024). Benefits and challenges of collaboration between students and conversational generative artificial intelligence in programming learning: An empirical case study. *Education Sciences*, 14(4), 433.

 © ISLS

International Society of the Learning Sciences

Fostering Cocreativity in Primary Education Through Viacocrea: Effects on Students' Creative Products

Maria Sitjà Martín, Manoli Pifarré Turmo
maria.sitja@udl.cat, manoli.pifarre@udl.cat
Universitat de Lleida

Abstract: This study explores how the Viacocrea technology platform enhances cocreative processes in primary education. Twenty-five twelve-year-old students participated in a project-based learning intervention, creating digital infographics collaboratively. Students' creative outcomes were assessed using Guilford's creativity dimensions (1950): fluency, flexibility, and originality. Findings show that Viacocrea effectively supported idea generation, adaptation, and originality, embedding creative processes throughout the project. This study demonstrates that technology can foster cocreativity within the curriculum, offering practical insights for teaching collaborative and creative skills in primary classrooms.

Introduction

Creativity is recognised as a key educational goal, fostering cognitive flexibility, originality, and innovative thinking (Cremin & Chappell, 2019). From a sociocultural perspective, it is a collaborative process shaped by dialogue and interaction (Sawyer, 2021; Glăveanu, 2010). In this view, cocreativity emphasises the joint generation of ideas and knowledge, going beyond individual contributions.

Digital technologies have increasingly been used to scaffold creative collaboration by supporting idea sharing, visualization, and iterative design processes in learning environments. Viacocrea addresses this gap by providing a shared, multi-user platform that fosters engagement and collaboration on real-world challenges. Through graphical representations of cocreative processes and a comprehensive toolkit, Viacocrea enables students to generate, refine, and combine ideas effectively, promoting iterative thinking and problem-solving (Pifarré, 2019).

The platform structures cocreation into six interconnected phases, such as Start up, Define, Design, Build Up, Sum up, and Communicate, supported by visual techniques and guidance for both divergent (e.g., brainstorming, idea elaboration) and convergent thinking (e.g., idea selection, reasoning). Viacocrea also systematically models a cocreative language, encouraging iterative thinking, questioning, exploration of perspectives, and reflection, which collectively cultivate collaborative and original outcomes.

This study aims to examine how a technology-enhanced cocreative pedagogy supported by Viacocrea affects primary students' creative processes and outcomes, specifically investigating its impact on the fluency, flexibility, and originality of their final products.

Methodology

This research followed a case study methodology in a sixth-grade classroom in Lleida (Spain) with 25 students (aged 12). Over the course of nine weeks, groups of 4-5 students created digital infographics to present aspects of their school and culture to peers in the United States. The project adopted a project-based learning approach, combining inquiry, collaboration, and creative use of technology.

Viacocrea supported the process by enabling shared brainstorming, refining, and consolidating ideas. Final products were analysed using a rubric aligned with Guilford's creativity dimensions (1950): fluency, flexibility, and originality (see Table 1).

Data collection

Students' final products were collected and assessed with a creativity rubric. These products took the form of digital infographics created collaboratively using Canva, an online multi-user design platform for creating visual content.

Data analysis

The final infographics were evaluated using a rubric based on Guilford's (1950) creativity dimensions: fluency, flexibility, and originality. Each dimension was operationalised into four criteria (see Table 1). Each criterion captured observable characteristics of the final infographic (e.g., diversity of ideas, inclusion of multiple perspectives, or novelty of content and design), allowing the creativity dimensions to be systematically assessed.

 © ISLS

Each criterion was scored on a 0-2 scale (0 = not observable; 1 = emerging; 2 = effective). Thus, each dimension had a maximum of 8 points, and total group scores could reach 24. Inter-rater agreement was established through independent scoring followed by discussion of discrepancies.

Table 1
Creativity Rubric Based on Guilford's Framework

Core creativity skills.	Specific criteria
Fluency	1. Clarity: The extent to which the information is presented clearly and understandably. 2. Organization: The clear and logical arrangement of elements within the infographic. 3. Coherence: Consistency between visual elements and the information presented. 4. Communication: Whether the infographic effectively conveys the intended message.
Flexibility	1. Adaptability: The infographic's capacity to adapt to different situations or audiences. 2. Multiple perspectives: Consideration of diverse viewpoints in the presented information. 3. Versatility: The infographic can show information in different ways. 4. Creative adaptation: Evidence of feedback incorporation to improve the infographic.
Originality	1. Unique approach: Distinctive perspective on the topic. 2. Innovative design: Originality in the infographic's visual and design elements. 3. Novel content: Inclusion of unconventional information or fun facts. 4. Uncommon interpretation: Presentation of ideas in an unconventional or visually appealing manner.

Findings and discussion

Analysis of the six final infographics indicated that most groups achieved high scores in fluency (clarity, coherence, effective communication) and flexibility (adaptability, multiple perspectives, creative adaptation). Several groups also demonstrated originality, producing distinctive perspectives and innovative designs.

However, variability across groups was observed, with one group scoring consistently lower. This suggests that while Viacocrea supports cocreative outcomes, its impact may depend on factors such as group dynamics, engagement, and prior experience. Further research should confirm this hypothesis. In addition, these findings underscore the importance of implementing targeted scaffolding strategies to support groups with lower participation levels and foster more equitable and inclusive cocreative processes. Table 2 presents the overall scores for each group.

Table 2
Overall Scores of Students' Final Products

	Fluency	Flexibility	Originality	Total Score
Group 1	8	8	8	24
Group 2	7	7	5	19
Group 3	5	6	6	17
Group 4	8	7	7	22
Group 5	4	4	4	12
Group 6	6	7	6	19

Overall, the results indicate that the platform helps structure cocreative processes by supporting idea generation (fluency), perspective-taking (flexibility), and the production of novel outputs (originality). These findings align with sociocultural perspectives that conceptualise creativity as a collaborative process emerging through interaction and shared artefacts (Sawyer, 2021; Glăveanu, 2010).

Conclusion

This study supports the hypothesis that a well-designed technology, such as Viacocrea, can foster cocreativity in primary education through project-based learning. By providing a shared space and structured creative toolkit, the platform scaffolded students to generate, adapt, and refine ideas collaboratively.

This study is limited by its small, single-class sample and variability across groups. Future research should adopt comparative and longitudinal designs to explore how technological scaffolding can sustain cocreative engagement and effectively support less active groups and provide a deep understanding of other educative variables that also play a role in cocreation.

 © ISLS

References

Cremin, T., & Chappell, K. (2019). Creative pedagogies: a systematic review. *Research Papers in Education, 36*(3), 1–33. https://doi.org/10.1080/02671522.2019.1677757

Glăveanu, V. P. (2010). Paradigms in the study of creativity: Introducing the perspective of cultural psychology. New Ideas in Psychology, *28*(1), 79–93.

Guilford, J. P. (1950). Creativity. *American Psychologist, 5*(9), 444–454. https://doi.org/10.1037/h0063487

Pifarré, M. (2019). Using interactive technologies to promote a dialogic space for creating collaboratively: A study in secondary education. *Thinking skills and creativity*, *32*, 1-16

Sawyer, R. K. (2021). The Dialogue of Creativity: Teaching the Creative Process by Animating Student Work as a Collaborating Creative Agent. Cognition and Instruction, 1–28.

Acknowledgement

This research has been funded by the Ministry of Science and Innovation of the Government of Spain under Grant number PID2022-139060OB-I00.

International Society of the Learning Sciences

AI as a Collaborative Partner: EMPA's Role in Supporting Reflective Teamwork Learning

Aparajita Jaiswal, Ashish Ashish
jaiswal2@purdue.edu, ashish@purdue.edu
Purdue University, West Lafayette, IN, USA

Abstract: Teamwork requires students to reflect on their practices, yet individualized mentoring is often limited. While AI tutoring systems increasingly support learning, most position AI as external evaluator rather than collaborative partner. We present EMPA, an AI-based collaborative partner that scaffolds teamwork learning through case studies, activities, and videos. Grounded in Schön's reflection-in-action and reflection-on-action, EMPA encourages students to adapt strategies during tasks and analyze experiences afterwards. Students engage with scenarios and activities while EMPA prompts real-time reflection, then receive personalized summaries and complete surveys that consolidate learning. Early deployment with undergraduates shows EMPA's potential to help learners examine assumptions, recognize cultural dynamics, and develop strategies for improving collaboration. This study operationalizes Schön's framework via AI-supported mentoring, positioning AI as a reflective partner to scale practice and strengthen teamwork.

Introduction

Teamwork is a cornerstone of higher education because it prepares students to solve complex problems and collaborate across diverse perspectives (Sridharan et al., 2023). Yet students often struggle with communication breakdowns, unequal participation, and cultural misunderstandings (Hussein, 2021). Although instructors can support collaboration through mentoring, such guidance is difficult to provide consistently in large classes (Navisotschnig, 2024). Recent advances in AI offer new possibilities. Rather than serving only as tools, AI systems can function as collaborative partners that engage learners through prompts, feedback, and reflective dialogue (Ashish et al., 2025). In this paper, we introduce EMPA, an AI-based collaborative partner grounded in Schön's reflective practice framework. Our research question is: *How can an AI collaborative partner operationalize Schön's reflective practice framework to improve teamwork?* EMPA supports both in-activity interpretation and retrospective reflection to help students develop adaptive teamwork strategies.

Theoretical grounding and operationalization of Schön's framework

The Schön's theory of reflective practice emphasizes reflection as central to professional learning in complex and uncertain contexts (Schön, 1991). He distinguishes between reflection-in-action, which occurs as learners adapt strategies during an activity, and reflection-on-action, which occurs afterwards as they analyze actions, outcomes, and assumptions. Teamwork provides an important domain for applying this framework because group work often involves conflict, miscommunication, and differing cultural norms (Jaiswal et al, 2024). Reflection-in-action helps learners respond flexibly in the moment, while reflection-on-action supports evaluation and future improvement. Prior research shows that scaffolding reflection can deepen collaboration, enhance metacognitive skills, and improve learning outcomes (Coulson & Harvey, 2013; Jia et al., 2025). However, these supports often remain external rather than participatory. EMPA extends this work by positioning AI as a collaborative partner in reflective cycles. EMPA operationalizes Schön's framework through interactional sequencing: reflection-in-action occurs during scenario analysis through real-time prompts, while reflection-on-action occurs after interaction through structured synthesis and personal strategy planning. Reflection-in-action is supported through real-time prompts embedded during engagement with case materials. As learners analyze unfolding teamwork scenarios, EMPA poses interpretive and metacognitive questions (e.g., "What assumptions might be shaping your interpretation?" or "How might this interaction be perceived differently across cultural contexts?"). These prompts intervene during engagement to support in-situ interpretation, prompting learners to reconsider assumptions and adapt reasoning while the scenario is unfolding. Reflection-on-action is supported through structured post-activity dialogue. Following scenario engagement, EMPA shifts roles to facilitate retrospective analysis focused on how learners would respond in their own teamwork contexts (e.g., "How would you respond differently in your own team?" or "What strategies would you carry into future collaborations?"). Here, EMPA shifts from moment-to-moment prompting to facilitating consolidation, abstraction, and transfer. Importantly, this distinction is operationalized through interaction timing, prompt structure, and conversational roles rather than through separate

 © ISLS

model training pipelines. The same LLM supports both reflective modes, but its behavior is shaped through system prompts, activity sequencing, and interface design that differentiate in-situ reflection from retrospective synthesis.

System architecture overview

EMPA combines an efficient LLM, modular interface, and scalable backend to support AI-mediated teamwork reflection. It is built on fine-tuned LLaMA-3 3B Instruct using QLoRA, enabling lightweight deployment while preserving performance. The browser-based interface (React/Tailwind, FastAPI, Supabase) guides learners through six modules with videos, quizzes, reflection prompts, and chat-based dialogue. FastAPI manages registration, conversations, and chat history, while Supabase stores user profiles and logs. Together, these components allow EMPA to function as a collaborative mentor that scaffolds reflection-in-action and reflection-on-action and helps learners transfer teamwork strategies to future contexts.

Pilot study and preliminary findings

To explore EMPA's potential as a collaborative partner, we conducted an initial pilot study with five undergraduate students. The study was designed to investigate how students engage with an AI partner during a reflective learning activity focused on teamwork. Five undergraduate students were recruited from an undergraduate course at a large public university. Participants engaged in a single, unmoderated session with the EMPA system. They were instructed to watch an embedded case study video depicting a multicultural project team facing common collaborative challenges. After the video, they engaged in a text-based dialogue with EMPA to interpret the team's dynamics and translate those observations into strategies for their own teamwork. We collected the full text logs of the conversations between each participant and EMPA for qualitative analysis. Our analysis of the conversation logs revealed that EMPA effectively scaffolds both of Schön's core reflective practices, positioning the AI as a collaborative partner in the learning process. ***Finding 1: Reflection-in-Action:*** Students engaged in reflection-in-action by analyzing unfolding team behaviors while interacting with EMPA. One student noted, *"The team dynamic broke down by Jonas interrupting the others and not letting them finish their ideas."* EMPA responded by contextualizing the comment: *"Some cultures value directness and assertiveness overactive listening. Reflecting on these potential biases can add depth to your analysis."* This exchange illustrates EMPA's ability to prompt reflection-in-action, encouraging learners to question assumptions as they engaged with the scenario. ***Finding 2: Scaffolding Collaboration Strategies:*** Students frequently proposed ways to improve the team dynamics within the case study. For example, one suggested, *"I would have told Jonas (character name) to wait a second and to hear the others out."* EMPA affirmed this and extended the idea: *"Consider implementing a round-robin approach, where each participant has an equal opportunity to contribute."* Similarly, another student concluded, *"If they gave each other time to hear each other out it would have worked better."* EMPA reinforced this by introducing strategies such as *active paraphrasing* and the *Stop-Think-Share* technique. These exchanges illustrate EMPA's role as a collaborative partner, supporting learners in co-constructing strategies they could apply within their own teamwork contexts. ***Finding 3: Fostering Reflection-on-Action:*** After the interaction, EMPA prompted reflection-on-action, guiding learners to reinterpret the scenario and formulate strategies for their own future teamwork. This is exemplified when a student proposed a sophisticated, actionable plan for future teamwork: *"I'd start with an empathy check: ask teammates early how they prefer to plan and what motivates them. I'd explicitly discuss expectations around deadlines and collaboration. By agreeing on a shared process at the start, we could prevent misunderstandings and respect everyone's working style."* This demonstrates a clear instance of reflection-on-action. Crucially, EMPA acts as a collaborative partner by not just validating the response, but by connecting it to theoretical concepts: *"Your proposed strategy is a great starting point. Your focus on active listening and prioritizing teamwork values aligns with collectivist cultures that emphasize harmony and cooperation."* By validating the student's idea (*"great starting point")* and co-constructing a deeper understanding by linking it to a theoretical concept (*"collectivist cultures"),* the AI moves beyond the role of a simple feedback tool into that of a true partner in learning. To conclude, this study introduced EMPA, an AI collaborative partner designed to operationalize Schön's reflective practice framework. Its main contribution is a proof-of-concept design that shifts AI from external evaluator to active participant in reflection. Although pilot findings (N=5) do not support claims of improved teamwork, they illustrate how EMPA's architecture and interaction sequencing can scaffold reflection-in-action and reflection-on-action. These formative results provide a foundation for larger studies of teamwork outcomes and adaptive mentoring.

References

Ashish, Jaiswal, A., Vhaduri, S., Nerella, N., & Jha, S. (2025). *Empa: An AI-powered virtual mentor for developing global collaboration skills in HPC education* [Preprint]. arXiv. https://arxiv.org/abs/2511.17669

Coulson, D., & Harvey, M. (2013). Scaffolding student reflection for experience-based learning: A framework. *Teaching in Higher Education*, *18*(4), 401-413.

Hussein, B. (2021). Addressing collaboration challenges in project-based learning: The student's perspective. *Education Sciences*, *11*(8), 434

Jaiswal, A., Sapkota, M., Ramos, M. M., Starr, L., & Stahl, A. (2024). Nurturing Teamwork, Intercultural Competence, and Belonging in First-Year STEM Students. *International Journal for the Scholarship of Teaching and Learning*, *18*(2), 4.

Jia, J., Ghazali, N. E., Ming, E. S. L., & Addi, M. M. (2025). The Impact of Scaffolding on the Development of Metacognitive Skills in Project-Based Engineering Learning. *Asean Journal of Engineering Education*, *9*(1), 1-10.

Navisotschnig, G. (2024). Teaching and Learning in Higher Education Intercultural Classrooms: Educators' Perspective on Classroom Challenges & Solutions Towards Inclusive Practice.

Schon, D. 1991. *The Reflective Practitioner*. 2nd edn. San Francisco: Jossey Bass.

Sridharan, B., McKay, J., & Boud, D. (2023). The four pillars of peer assessment for collaborative teamwork in higher education. In *The Power of Peer Learning: Fostering Students' Learning Processes and Outcomes* (pp. 3-24). Cham: Springer International Publishing.

Acknowledgments

This work was supported by the Innovation Hub under grant number IH-AI-24012. We gratefully acknowledge their support in enabling the development of EMPA.

International Society of the Learning Sciences

Boundary Spanning Narratives in a Constructionist Museum Game

Eda Zhang, University of Wisconsin–Madison, rzhang346@wisc.edu
Matthew Berland, University of Wisconsin–Madison, mberland@wisc.edu
Yunseo Lee, University of Wisconsin-Madison, ylee899@wisc.edu

Abstract: Museum learning is highly social (Allen, 2004). To support and enable socioculturally mediated learning, previous studies have demonstrated how designing exhibits for collaboration can yield productive learning outcomes and engagement. However, few studies have explored the learning value of naturalistic collaborative play in museums (in which there are no explicit prompts/scaffolds for collaboration). We present a museum game and both computational and qualitative observations of interaction modes. Visitors showed diverse collaborative and cooperative behaviors, including clarification, seeking help, negotiation, goal-adaptation, and modeling. Our analysis also revealed that boundary spanning narration (or BSN, Tissenbaum et al., 2017) enabled "talk with evidence of learning." Additionally, we found that social play (in this case, cooperative and competitive play) varied with the types and amount of BSN. Through the mediation of BSN, two-person visitor groups ("dyads") co-constructed knowledge, coordinated on tasks, and set shared goals for themselves in non-competitive play, while in competitive play, dyads tended to use BSN towards surveillance and dominating play.

Introduction and theoretical framework

Museums are loci of public connections that allow diverse audiences to join and thus have great potential in building diverse communities of learning (Falk, 2002; Horrigan, 2016; Berland, 2020; Wardrip et al., 2017). Learning in museums is socially situated (Allen, 2002; Falk & Dierking, 2016), so, in order to understand how people learn with others and build a community of creating, we need to understand how visitors learn through interacting with each other in naturalistic settings. Prior research has primarily examined collaborative learning through explicitly designed structures, such as assigned roles (Klopfer & Squire, 2005), shared tasks (Apostolellis & Bowman, 2015), or challenges requiring joint action (Rowe et al., 2014). However, we know far less about the naturalistic social interactions that emerge between youths when playing constructionist museum games. Therefore, our research question is: **how do youth visitors play differently when a constructionist museum game is played competitively versus when playing collaboratively**?

We investigate the question through *Rainbow Agents*, a 2-player constructionist museum game which is designed to support both individual and group learning of computational literacy. The focus of the game is to invite players to build their programs to tend a shared garden with animal characters that are agents that execute the coding logics they created. The game does not explicitly prompt for tight-coupled collaboration, instead, it just invites players to tend the garden, which enables naturalistic social interaction to emerge. Through this work, we explore the learning engagement and the collaborative learning mechanism that emerges from naturalistic social interactions in museums. In the game, players program animal characters to tend a shared garden using computational logic cards. Each player operates a touchscreen console for character selection and programming, while a large TV displays the communal garden space. The game features three programmable creatures, each embodying distinct computational modeling concepts (e.g., loops, conditionals, sequences). Players select territories (3×3 tiles), choose cards, and sequence actions such as planting and watering. While the game-directed goal involves filling three orbs to trigger a storm or open a treasure chest, the constructionist design encourages emergent, player-defined objectives. The game does not mandate tight-coupled collaboration, allowing naturalistic social interactions to unfold organically.

Figure 1
Rainbow Agents



While several studies focus on explicit collaboration in museums, less attention has been paid to how collaborative learning mechanisms might function differently across various interaction modes, including competitive play. So far, mixed results on the effect of competition on game-based learning have been reported. Chen et al. (2020) suggests that competition was often more effective for math, science, and language and less so social science and other subjects. Plass et al. (2013) found that competition boosted motivation and within-game

 © ISLS

performance in a math fluency game compared to individual play. But out-of-game fluency gains occurred across all groups regardless of play modes. Vandercruysse et al.(2013) found that players focused on the scores while engaged less in learning materials and attribute more value to winning rather than the task. Even less has been studied in which ways people communicate and what learning looks like in competitive play, and what it looks like in museums. Multiplayer games offer a productive lens for examining the relationship between informal collaborative and competitive play (Gee, 2008; Plass et al., 2015; Steinkuehler, 2012). Visitors come to museums with diverse backgrounds and divergent personal goals, raising questions about how collaborative learning emerges when participants are not working toward a shared objective. Tissenbaum et al. (2017) demonstrated that museum visitors can engage in productive collaboration even with divergent goals, introducing the Divergent Collaborative Learning Mechanisms (DCLM) framework to capture these interactions. Building on this work, we examine how collaborative learning mechanisms emerge across different interaction modes including cooperative, collaborative, and competitive play in a two-player constructionist game designed to develop computational literacy.

Method

Data was collected at a science and technology museum in North America. At the exhibit, we used a "lollipop" consent sign to indicate participation and permission for video/audio recording. Over 10 hours of video were recorded using a fixed camera positioned behind the exhibit to capture gestures and gaze without disrupting visitors. Audio was captured via table microphones, and system interactions were logged automatically for all consented participants. Our study employed interaction analysis (Jordan & Henderson, 1995; Ash, 2003) within a multiple case study design (Yin, 2009) to investigate how learning processes differ across interaction modes in museum constructionist gaming.

Results, discussion, and conclusions

We found that competitive play and non-competitive play varied significantly in learning process and learning engagement. In competitive play, the narration is dominated by dominance claiming and resources bragging. The effects of competition in game-based learning are mixed. Results from our study showed that players' competition tends to limit individual as well as collective mean-making. We observed players engaged overwhelmingly in the idea of "winning/conquering" over the actual game content (Vandercruysse et al., 2013), thus engage heavily in score talk (in our case, territory dominance claiming), echoing the results of Meece et al. (2005). We found that their talk is qualitatively different in competitive play vs. cooperative play. In competitive play, the talk was unidirectional rather than reciprocal, showing a misaligned engagement on topics and interests, which may have inhibited the collaborative interactions. Similar to Kohn (1992)'s finding, we saw player anxiety and frustration in this environment; players more likely turned to a competitive play mode.

Tissenbaum et al. (2017) showed that learners who are simultaneously engaged in divergent (or non-convergent) goals can still effectively scaffold each other and engage in productive collaboration. In this study, we provide further empirical evidence for these findings. Additionally, we found players engage in fluid and dynamic interaction modes, switching from individual/parallel play to cooperative or collaborative play, and generate productive collaborative learning interactions even when no explicit tight collaboration is prompted. Our analysis revealed that a key mechanism that channeled players' activities to collaborative learning is Boundary Spanning Narration (BSN). In *Rainbow Agents*, it is designed to be easy to observe the other player's activities along with themselves because of the shared display and side-by-side seating design. The shared display shows the shared communal garden space that belongs to both players, which allows them to watch things about themselves, about the other player, as well as the overview of the shared group status. Previous study about multi-user tabletop territoriality has demonstrated that people categorize the space into personal and group uses in using shared space with others (Scott et al., 2004). In this game, there is neither a single dedicated personal space, nor any "boundaries" of group and personal space. The ambiguity of ownership of the territory may have led to the behavior of boundary spanning perception and other forms of narration, thus incubating multiple BSN-facilitated collaborative interactions (e.g., scaffolding, meaning-making, assess and review).

In this paper, we investigated the learning of competitive and cooperative play in a 2-player constructionist museum game. Our results showed limited individual and collective meaning-making in the competitive play, while visitors engaged in a variety of collaborative learning activities through boundary spanning narration (BSN) mechanism. This study is an exploratory one by nature as we used only a small part of a large data set. In the future, quantitative analysis on BSN can reveal scalable insights about the effect of BSN in terms of game-directed learning outcomes. Additionally, future design can consider using a shared display to encourage fluid and dynamic interaction modes between users.

 © ISLS

References

Allen, S. (2003). Looking for learning in visitor talk: A methodological exploration. In Learning conversations in museums (pp. 265-309). Routledge.

Apostolellis, P., & Bowman, D. A. (2015, June). Small group learning with games in museums: effects of interactivity as mediated by cultural differences. In *Proceedings of the 14th International Conference on Interaction Design and Children* (pp. 160-169).

Ash, D. (2003). Dialogic inquiry in life science conversations of family groups in a museum. *Journal of Research in Science teaching*, *40*(2), 138-162.

Berland, M. (2020). Playful constructionism in museums, libraries, and other voluntary spaces: co-construction, co-facilitation, and co-dread. Designing constructionist futures, 205-212.

Chen, C. H., Shih, C. C., & Law, V. (2020). The effects of competition in digital game-based learning (DGBL): a meta-analysis. *Educational Technology Research and Development*, *68*(4), 1855-1873.

Crowley, K., & Callanan, M. (1998). Describing and supporting collaborative scientific thinking in parent-child interactions. *Journal of Museum Education*, *23*(1), 12-17.

Deutsch, M. (1949). A theory of co-operation and competition. *Human relations*, *2*(2), 129-152.

Dillenbourg, P. (1999). What do you mean by collaborative learning?. Collaborative-learning: Cognitive and computational approaches., 1-19.

Falk, J. H. (2002). The contribution of free-choice learning to public understanding of science. Interciencia, *27*(2), 62-65.

Gee, J. P. (2008). A sociocultural perspective on opportunity to learn. Assessment, equity, and opportunity to learn, 76-108.

Vom Lehn, D., Heath, C., & Hindmarsh, J. (2001). Exhibiting interaction: Conduct and collaboration in museums and galleries. *Symbolic interaction*, *24*(2), 189-216.

Jordan, B., & Henderson, A. (1995). Interaction analysis: Foundations and practice. *The journal of the learning sciences*, *4*(1), 39-103.

Klopfer, E., & Squire, K. (2008). Environmental Detectives—the development of an augmented reality platform for environmental simulations. *Educational technology research and development*, *56*(2), 203-228.

Kohn, A. (1992). No contest: The case against competition. Houghton Mifflin Harcourt.

Meece, J. L., Anderman, E. M., & Anderman, L. H. (2006). Classroom goal structure, student motivation, and academic achievement. *Annu. Rev. Psychol.*, *57*(1), 487-503.

Pellicone, A., Lyons, L., Kumar, V., Zhang, E., & Berland, M. (2019, October). Rainbow agents: A collaborative game for computational literacy. In *Extended abstracts of the annual symposium on computer-human interaction in play companion extended abstracts* (pp. 597-604).

Plass, J. L., O'Keefe, P. A., Homer, B. D., Case, J., Hayward, E. O., Stein, M., & Perlin, K. (2013). The impact of individual, competitive, and collaborative mathematics game play on learning, performance, and motivation. *Journal of educational psychology*, *105*(4), 1050.

Rowe, J. P., Lobene, E. V., Mott, B. W., & Lester, J. C. (2014, January). Play in the museum: Designing game-based learning environments for informal education settings. In FDG.

Scott, S. D., Carpendale, M. S. T., & Inkpen, K. (2004, November). Territoriality in collaborative tabletop workspaces. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work* (pp. 294-303).

Steinkuehler, C. A. (2012). Learning in massively multiplayer online games. In Embracing Diversity in the Learning Sciences (pp. 522-529). Routledge.

Tissenbaum, M., Berland, M., & Lyons, L. (2017). DCLM framework: understanding collaboration in open-ended tabletop learning environments. *International Journal of Computer-Supported Collaborative Learning*, *12*(1), 35-64.

Vandercruysse, S., Vandewaetere, M., Cornillie, F., & Clarebout, G. (2013). Competition and students' perceptions in a game-based language learning environment. *Educational Technology Research and Development*, *61*(6), 927-950.

Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes (Vol. 86). Harvard university press.

Wardrip, P., Brahms, L., Reich, C., & Carrigan, T. (2017). Making+ learning in museums and libraries. A practitioner's guide and framework. *Children's Museum of Pittsburgh and the Institute of Museum and Library Services.* Retrieved from https://makingandlearning. squarespace. com/publication/Most have a deadline of Dec, 1.

Yin, R. K. (2009). Case study research: Design and methods (Vol. 5). sage.

International Society of the Learning Sciences

Neurodiversity in Computer-Supported Collaborative Learning: A Literature Review

Greta Heathcote, Alison Nelson, Armando Paulino Preciado Babb
greta.heathcote1@ucalgary.ca, alison.penn@ucalgary.ca, apprecia@ucalgary.ca
University of Calgary

Abstract: Recent evidence has shown a significant prevalence of neurodivergence in the general population, suggesting that many learners, including students and research participants, are neurodivergent. This situation entails potential concerns for both research results interpretations and ethical considerations. In this context, we explored the intersection of neurodiversity and the CSCL through a review of its international journal and handbook, finding a minimal use of neurodiversity-related terminology, despite indirect references to neurodivergent traits. The study identified 58 publications with potential reference to the neurodivergent population by filtering key terms—such as equity, inclusion, and low ability—that align with neurodivergent experiences, though often without explicit acknowledgment. We conclude that there is a significant gap in explicit consideration of neurodivergent learners within CSCL scholarship and argue that the CSCL community may have inadvertently included neurodivergent participants.

Introduction

Insights on neurodiversity have permeated through education research and practice, with respect to neurodivergent people. Broadly speaking, the term neurodiversity refers to natural variation of differences in human brain functions and behaviours, while neurodivergent refers to those whose natural variances impede their social daily lives or their ability to participate in classroom learning. The neurodevelopmental conditions, such as Attention Deficit Hyperactivity Disorder, Autism Spectrum Disorder, and other learning disabilities are considered under the umbrella of neurodivergent variations affecting cognitive, affectual, and sensory function (Bertilsdotter Rosqvist et al., 2020). Some studies estimate that between 15% and 20% of the student population is included in the neurodiversity label (Hamilton & Petty, 2023; Statistics Canada, 2023).

Such a high prevalence of neurodiverse students represents both an ethical concern and an opportunity for the Computer-Supported Collaborative Learning (CSCL) community. Studies within CSCL might have inadvertently involved neurodivergent participants, with potential implications for the conclusions drawn from research findings. The research team consists of two doctoral students, including the lead author who identifies as neurodivergent, and an Associate Professor who is committed to advancing neuro-inclusion and awareness in CSCL. We decided to review the CSCL literature, from the international journal (ijCSCL) and the international handbook (iHCSCL), regarding the potential involvement of neurodivergent population. While we recognize that ijCSCL and iHCSCL show more prevalence towards typical learners, we decided to investigate potential evidence of a neurodivergent population involved within CSCL scholarship. Advocating for the consideration of the neurodivergent population by the CSCL community, we report the results of a literature review exploring how ijCSCL and iHCSCL may have involved neurodivergent individuals and provide recommendations for considering neurodiversity in educational research and practice pertaining to CSCL.

Methodology

We searched publications from the ijCSCL dated 2006-2024 (n=457) and iHCSCL (n=35). We filtered through the titles and abstracts using terms such as ‘neurodiversity,’ ‘neurodivergent,’ ‘special education,’ and ‘accessibility’ to identify references to the neurodivergent movement. These search terms yielded no results, and we decided to conduct a systematic review using terms often associated with the neurodivergent population. We followed a multiplier approach (Page et al., 2021) for systematic reviews regarding selecting and filtering relevant articles. For inclusion criteria, we focused on all the publications up to December 2024 in ijCSCL and iHCSCL.We refined our sources from ijCSCL database by removing duplicates (n=395), corrections (n=387), erratums (n=385), and book reviews (n=384) before screening articles that met our selection criteria.

The terms we used for searching documents related to the neurodivergent population derived from the literature and our experience in the classroom and educational institutions: “bias,” “collaboration,” “equity,” “intervention,” “low,” “regulation,” “risk,” and “visual.” We identified 63 publications using at least one of these terms in the title or the abstract. Then, we reviewed the publications for relevance to our interest, selecting 58 documents for the review.

 © ISLS

Findings

Our findings are revealed through a timeline dated 2006-2024 and representation of terms. The most noteworthy years were 2006 (n=9), 2008 (n=7) and 2011 (n=12) as they had larger numbers of articles published that contained, defined and contextualized terms aligning with the neurodivergent movement. It was in 2008 when, for the first time, the terms 'inclusion' and 'bias' offered contextual understanding and illuminated barriers that align with the neurodiversity movement. In 2011, we found publications including the term 'equity' and offering a contextual understanding and illuminated barriers that align with group participation.

Representation of terms

The most noteworthy terms were 'collaboration' (n=58), 'low' (n=12), and 'equity' (n=5). It was probable that 'collaboration' would have the highest term representation due to the focus of CSCL. 'Collaboration' has the highest term representation when other terms such as 'equity' and 'inclusion' are the lowest term representation. This is because key components of collaboration are equity and inclusion; therefore, it would have been hypothesized that equity, inclusion, and collaboration would have had similar term representation.

The term 'low' was filtered generally as many researchers described participants or their level of understanding in different contents. For example, researchers wrote 'lower order thinking skills' (Wegerif, 2006), 'low agreement ratings' (Buder & Bodemer, 2008), 'low prior knowledge' (Ke, 2008), 'low achievers' (Molenaar et al., 2011), 'low ability' (Molenaar et al., 2011; Walker et al., 2011), and 'low-level argumentation' (Pontual Falcão & Price, 2011) however it was not until 2011 that Cress et al. (2011) wrote 'intellectual disability.' Using and contextualizing 'intellectual disability' aligns more with the neurodivergent movement as 'low-ability' or 'low-achievers' can relate to multiple levels of learners.

Equity was first offered contextual understanding in 2011. Schwarz et al. (2018) recognized the difference between equity and equality in their research. They researched the emergence of conceptual learning in a collaborative setting, elaborating on critical moments in group learning and teachers' identification of critical moments to scaffold students to the emergence of conceptual learning (Schwarz et al., 2018). This distinction between equity and equality is important to meet the individual needs of student wellbeing and learning goals. However, the term 'equity' was not represented again in the selected and filtered articles after this publication until the iHCSCL published a chapter interrogating the role of equity, diversity and inclusion in CSCL (Gomez, et al., 2021). In this investigation, Gomez et al. (2021) define Equity, Diversity and Inclusion (EDI) within the context of CSCL and outline the challenges of scalability and identify three critical issues: language, differentiation, and identity.

Discussion

Our discussion on the terminology within the ijCSCL and iHCSCL relates to the rise of the neurodiversity movement that has broadened over time. Neurodiversity is not a diagnosis but an identity that encompasses a variety of neurodevelopmental and more recently, mental health conditions and seeks to move the conversation away from medical 'disorders' and instead focusing on how an individual perceives the world (Bertilsdotter Rosqvist et al., 2020). This context is important as the language changes as our understanding of neurodevelopmental conditions increases and research expands, we need to look through early research in ijCSCL and iHCSCL with the lens of behaviours, characteristics, that may represent neurodivergent traits. As we considered terms that might assist us in finding connections to supporting neurodivergent learning, we selected terms that are frequently used in discussions of supporting learners with social, cognitive, academic, and wellbeing needs. With these terms, we found evidence that research in CSCL may have included individuals with neurodivergent traits, which is expected due to the high prevalence of this population. Using such terms as a framework can help researchers situate neurodivergent learning within the literature that could otherwise seem disconnected or even disinterested in supporting equity-deserving students.

Conclusion

In this paper we elaborate on the evidence of the high prevalence of neurodivergent people, demonstrating that both the journal and the handbook present research and findings that connect to neurodivergent learners. The findings present a general landscape of the current state of knowledge for the ijCSCL and iHCSCL and its terminology. Using this terminology as a starting point to build a framework can help CSCL researchers situate neurodivergent learning within studies that could otherwise seem disconnected or even disinterested in supporting equity- deserving students whereas we argue that CSCL researchers are indeed creating important work connected to accessible and equitable learning environments.

References

Bertilsdotter Rosqvist, H., Stenning, A., & Chown, N. (2020). Introduction. In H. Rosqvist, N. Chown, & A. Stenning (Eds.), *Neurodiversity studies: A new critical paradigm* (pp. 1-11). Routledge. https://doi.org/10.4324/9780429322297-1

Buder, J., & Bodemer, D. (2008). Supporting controversial CSCL discussions with augmented group awareness tools. *International Journal of Computer-Supported Collaborative Learning*, *3*(2), 123–139. https://doi.org/10.1007/s11412-008-9037-5

Cress, U., Wodzicki, K., Bientzle, M., & Lingnau, A. (2011). CSCL for intellectually disabled pupils: Stimulating interaction by using a floor control mechanism. *International Journal of Computer-Supported Collaborative Learning*, *6*(2), 307–321. https://doi.org/10.1007/s11412-010-9105-5

Gomez, K., Gomez, L. M., & Worsley, M. (2021). Interrogating the role of CSCL in diversity, equity, and inclusion. U. Cress, C. Penstein Rosé, A Friend Wise, & J. Oshima (Eds.), In *International Handbook of Computer-Supported Collaborative Learning* (pp. 103–119). Springer International. https://10.1007/978-3-030-65291-3_6

Hamilton, L. G., & Petty, S. (2023). Compassionate pedagogy for neurodiversity in higher education: A conceptual analysis. *Frontiers in Psychology*, *14*, 1-9. https://doi.org/10.3389/fpsyg.2023.1093290/full

Ke, F. (2008). Alternative goal structures for computer game-based learning. *International Journal of Computer-Supported Collaborative Learning*, *3*(4), 429–445. https://doi.org/10.1007/s11412-008-9048-2

Martínez-Jaime, M. M., Reyes-Morales, H., Peyrot-Negrete, I., & Barrientos-Álvarez, M. S. (2024). Access to early diagnosis for attention-deficit/hyperactivity disorder among children and adolescents in Mexico City at specialized mental health services. *BMC Health Services Research*, *24*(1), 599.

Molenaar, I., Chiu, M. M., Sleegers, P., & van Boxtel, C. (2011). Scaffolding of small groups' metacognitive activities with an avatar. *International Journal of Computer-Supported Collaborative Learning*, *6*(4), 601–624. https://doi.org/10.1007/s11412-011-9130-z

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., . . . Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Systematic Reviews*, *10*(1), 89. https://doi.org/10.1186/s13643-021-01626-4

Pontual Falcão, T., & Price, S. (2011). Interfering and resolving: How tabletop interaction facilitates co-construction of argumentative knowledge. *International Journal of Computer-Supported Collaborative Learning*, *6*(4), 539–559. https://doi.org/10.1007/s11412-010-9101-9

Schwarz, B. B., Prusak, N., Swidan, O., Livny, A., Gal, K., & Segal, A. (2018). Orchestrating the emergence of conceptual learning: A case study in a geometry class. *International Journal of Computer-Supported Collaborative Learning*, *13*(2), 189–211. https://doi.org/10.1007/s11412-018-9276-z

Statistics Canada. (2023). *New data on disability in Canada, 2022*. https://www150.statcan.gc.ca/n1/pub/11-627-m/11-627-m2023063-eng.htm

Walker, E., Rummel, N., & Koedinger, K. R. (2011). Designing automated adaptive support to improve student helping behaviors in a peer tutoring activity. *International Journal of Computer-Supported Collaborative Learning*, *6*(2), 279–306. https://doi.org/10.1007/s11412-011-9111-2

Wegerif, R. (2006). A dialogic understanding of the relationship between CSCL and teaching thinking skills. *International Journal of Computer-Supported Collaborative Learning*, *1*(1), 143–157. https://doi.org/10.1007/s11412-006-6840-8

International Society of the Learning Sciences

What Draws Upvotes? Examining Relationships between Interpersonal Skills and Social Annotation Behaviors in Online Settings

Hanxiang Du, Western Washington University, duh2@wwu.edu
Jiyoung Cheong, Western Washington University, cheongj@wwu.edu
Gaoxia Zhu, Nanyang Technological University, gaoxia.zhu@nie.edu.sg
Xiaoshan Huang, McGill University, xiaoshan.huang@mail.mcgill.ca
Shan Li, Lehigh University, shla22@lehigh.edu
Juan Zheng, Lehigh University, juz322@lehigh.edu
Jacqueline Tan, Western Washington University, tanj3@wwu.edu
Tianlong Zhong, Nanyang Technological University, tianlong001@e.ntu.edu.sg
Chenyu Hou, Nanyang Technological University, chenyu004@e.ntu.edu.sg
Tongxi Liu, University of Macau, txliu@um.edu.mo

Abstract: In social annotation learning settings, upvotes serve as feedback and a source of motivation, social connection, and validation, yet factors influencing upvote patterns remain under-explored. This study examined relationships between upvote count and students' interpersonal skills, performance, and annotation behaviors. We analyzed 7,482 annotations from 93 undergraduate students using Perusall over seven weeks. We observed no strong correlations between the upvote count and students' interpersonal skills, automated evaluation score, reply count, or posting order.

Objectives and significance

Social annotation, which refers to a collaborative process in which instructors and students highlight, comment on, and discuss course materials (Novak et al., 2012), has become a transformative pedagogical approach to promote active learning and community-building in online learning. It fosters deeper reading engagement and comprehension by making thinking visible and creating opportunities for intellectual exchange (Zhu et al., 2020; Huang et al., 2024). Upvoting is a prevalent feature in social annotation platforms, enabling users to express approval and engage with peers' contributions (Li et al., 2024). Upvotes can serve as a powerful social signal that influences content visibility and shapes discourse. Previous studies indicate that upvotes in social annotation are associated with both content-related and interpersonal cues and can potentially shape students' learning experience and perceived learning outcomes (Huang et al., 2024). However, the empirical evidence about the factors behind these upvote count in online social annotation settings remains limited. The current study seeks to address this gap. Specifically, we focused on the following question: *What are the relationships between upvote count and students' interpersonal skills and social annotation behaviors?*

Study design and methods

This study was conducted during the Winter semester in an undergraduate learning media course at a North American university with 97 students (47 female; Age *Mean* = 20.67). Four students withdrew, yielding a final sample of 93. Students were assigned to 16 small groups and engaged in collaborative reading and discussion activities using Perusall, which allows students to asynchronously annotate course readings by highlighting text, posing texts, and replying. The platform also features automatic scoring that evaluates students' contributions based on criteria such as relevance, quality, and engagement. Over the seven-week activities, students were required to annotate ten readings. A total of 7,482 annotations were collected, including original annotations, replies, upvotes, timestamps, and system-generated scores.

To examine individuals' interpersonal competencies, we adapted an interpersonal skills survey based on Riggio's six dimensions (Lee et al., 2015). The survey consists of seven items measuring self-perceived behaviors such as empathy, open-mindedness, communication, conflict resolution, and adaptability, rated on a 7-point Likert scale (1 = Strongly Disagree, 7 = Strongly Agree). A total of 84 participants completed the survey.

To address the research questions, we examined the relationship among several key variables. Students' responses to the interpersonal skill surveys were used to measure their interpersonal skills. The content quality of their annotations was measured using Perusall's automated evaluation score. We also explored students' social annotation behaviors in terms of reply count and posting order. Reply count is the total number of replies an annotation received, and posting order refers to the time an annotation was posted. To identify the posting order,

 © ISLS

we retrieved the posting timestamp for each annotation within each set of reading materials, with the earliest social annotation marked as 1. In other words, if a document contains *n* annotations, the most recently posted annotation has the posting order *n*.

Results

To address the research question, we analyzed several relationships: (1) upvote count and the automated evaluation score provided by Perusall, (2) upvote count and the number of replies each annotation received, (3) upvote count and the posting order of each annotation, and (4) students' total upvote count and their interpersonal skill scores. Table 1 presents the results using Spearman correlation. Although the first three variables were statistically significant, all variables were weakly correlated with the upvote count. Posting order and interpersonal skills were negatively and weakly correlated with the upvote count.

Table 1

The Correlation between the Upvote Count and Automated Evaluation Score, Online Social Annotation Behaviors, and Interpersonal Skills

	Automated Score	Reply Count	Posting Order	Interpersonal Skills
Upvote count	.107*	.159*	-.050*	-.027

*: $p < 0.05$

Discussion and conclusion

This study examined factors influencing upvote patterns in social annotation by analyzing 7,482 annotations from 93 undergraduate students. Contrary to expectations based on social presence theory (Arbaugh et al., 2008), interpersonal skills showed weak negative correlations with upvote counts. Students with higher self-reported interpersonal competencies did not receive more peer recognition through upvotes. This result challenges assumptions about social dynamics in online learning environments (Chen et al., 2024). Similarly, Perusall's automated evaluation scores correlated only weakly with upvotes. This disconnection between algorithmic assessment and peer recognition suggests that students value different qualities in annotations than those captured by automated systems. Peer preferences appear to diverge from traditional academic quality metrics. The weak relationship between automated scores and upvotes also suggests caution in using peer feedback mechanisms for assessment purposes. Upvotes appear to reflect student interests rather than content quality, making them unsuitable as primary evaluation tools.

There are limitations to this work. Our findings were based on certain topics in that course and the generalization to other disciplines may be limited. Future research should explore the cognitive and emotional factors underlying topic preferences in upvoting behavior. Understanding why certain topics consistently attract more peer acknowledgement could inform more equitable discussion design. Moreover, investigating student motivations for upvoting through interviews or structured feedback would provide valuable insights into peer recognition patterns. Finally, researchers might explore whether instructor intervention can balance attention across different annotation types.

References

Arbaugh, J. B., Cleveland-Innes, M., Diaz, S. R., Garrison, D. R., Ice, P., Richardson, J. C., & Swan, K. P. (2008). Developing a community of inquiry instrument: Testing a measure of the community of inquiry framework using a multi-institutional sample. *The Internet and Higher Education*, 11(3-4), 133-136.

Chen, F., Li, S., Lin, L., & Huang, X. (2024). Identifying temporal changes in student engagement in social annotation during online collaborative reading. *Education and Information Technologies*, 29(13), 16101-16124.

Huang, X., Zheng, J., Li, S., Zhu, G., Du, H., Zhong, T., ... & Lajoie, S. (2024). Investigating the effect of emotional tone on learners' reading engagement and peer acknowledgement in social annotation. *Australasian Journal of Educational Technology*, *40*(6), 92-107.

Lee, D., Huh, Y., & Reigeluth, C. M. (2015). Collaboration, intragroup conflict, and social skills in project-based learning. *Instructional science*, 43(5), 561-590.

Li, S., Huang, X., Lin, L., & Chen, F. (2024). Exploring influential factors in peer upvoting within social annotation. *British Journal of Educational Technology*, 55(5), 2212-2225.

Novak, E., Razzouk, R., & Johnson, T. E. (2012). The educational use of social annotation tools in higher education: A literature review. The Internet and Higher Education, 15(1), 39-49.

Zhu, X., Chen, B., Avadhanam, R. M., Shui, H., & Zhang, R. Z. (2020). Reading and connecting: Using social annotation in online classes. *Information and Learning Sciences*, 121(5/6), 261-271.

 © ISLS

International Society of the Learning Sciences

Exploring STEM Teachers' AI Readiness, Attitudes, and Uses of Generative AI across Teaching Subjects

Lin Yung-Hsuan, National Tsing Hua University, taro90085@gmail.com
Lin Pei-Yi, National Tsing Hua University, peiyilin@mx.nthu.edu.tw

Abstract: This study explores the integration of AI among Taiwanese STEM teachers, revealing a shift from utilitarian tool use toward co-creative partnership. While efficiency-driven tasks currently predominate, innovative applications are emerging in interdisciplinary contexts. Disciplinary norms significantly influence adoption: mathematics teachers prioritize precision, while science and interdisciplinary teachers leverage AI for creative integration. The findings suggest that professional development could provide discipline-specific support to empower teachers as collaborative partners in AI-integrated classrooms.

Introduction and theoretical framework

Teachers in schools are widely adopting artificial intelligence (AI) technologies to improve the quality and effectiveness of their work. They have been used in teaching to help teachers analyze student performance and provide immediate support (Esiyok et al., 2024). Recently, generative AI has been shown to aid teachers in lesson preparation, for example, generating relevant teaching materials (Kim et al., 2022). It is necessary to understand teachers' AI readiness and their experiences incorporating AI technologies into their teaching practices (Miao & Cukurova, 2024). Previous studies suggest that teachers with higher AI readiness may be more capable of deploying AI technologies to support their teaching (Luckin et al., 2022). However, Popenici and Kerr (2017) noted that teachers need to rethink AI and teachers' roles in schools. Teachers should have a comprehensive understanding of AI and know when, where, and how to apply it, along with AI-generated data, to optimize their teaching practices (Nazaretsky et al., 2022). Previous studies have used quantitative questionnaires to investigate teachers' readiness and behavioral intentions regarding AI use (e.g., Ramazanoglu & Akın, 2024; Sanusi et al., 2024). However, while large-scale quantitative studies can provide an overview of teachers' AI readiness, they fail to fully account for the actual use and various considerations that teachers across disciplines face when using AI to prepare for teaching tasks, such as the knowledge-creation and critical-thinking processes involved in adopting AI-generated knowledge. Therefore, this study aims to use semi-structured interviews to understand teachers' AI readiness, attitudes, and actual use of AI technologies. The research questions proposed in this study are: (1) What were the teachers' AI readiness, attitudes, and uses when using generative AI? (2) How do teachers' AI readiness, attitudes, and uses differ regarding teaching subjects?

Method

This study used purposive and convenience sampling methods. We recruited STEM teachers with experience teaching in northern Taiwan for interviews in the summer of 2025. Participant selection criteria included: (1) having at least one semester of K-12 teaching experience; and (2) having experience using generative AI in school administration or teaching work. Ten participants from K-12 public schools in northern Taiwan participated in this study. These included 3 mathematics teachers, 3 natural science teachers, 1 information technology teacher, and 3 interdisciplinary STEM teachers. They had taught for at least one semester in elementary (n = 5) and secondary (n = 5) education. The researchers adopted semi-structured interviews and open-ended questions. Participants completed an informed consent form and reviewed the interview questions in advance. Interview questions included teachers' background (e.g., age, years of teaching experience, university background, grade levels, and teaching subject), teachers' understanding of AI concepts, attitudes (e.g., How do you feel about the development of AI tools?), responsible AI use (e.g., How do you use AI tools to search for or generate knowledge?), concern about AI's development (e.g., What role do you think AI tools will play in future school education?). The interview was recorded and transcribed with the participant's consent.

The qualitative data collected from these interviews were analyzed using an open coding approach. The data were categorized into six themes: (1) Teachers' practical uses of AI: Theme 1-1 Instructional support, 1-2 Student learning, 1-3 Teachers' uses, 1-4 Limited use. (2) Teachers' understanding of AI: Theme 2-1 Structure information, 2-2 Generate diverse ideas, 2-3 Provide immediate feedback. (3) Teachers' perspectives towards AI: Theme 3-1 Tool, 3-2 Partner, 3-3 Context and experience dependence. (4) Teachers' emotional responses: Theme 4-1 Positive, 4-2 Negative, 4-3 Co-existing. (5) Risks of AI: Theme 5-1 Weaken students' cognition, 5-2 Content quality issue, 5-3 Behavioral problems. (6) Response to the uses of AI: Theme 6-1 Instructional adjustments, 6-2 Teacher professional judgment, 6-3 Information literacy.

 © ISLS

Primary findings

Teachers' attitudes toward AI: Between “tool” and “co-creative partner”

Interviews suggest that teachers are cautious yet optimistic. They recognize AI's usefulness in handling routine tasks but are hesitant to rely on it entirely. Some repeatedly stressed that “AI is just a tool” (Theme 3-1), primarily for administrative support (Theme 1-4), item generation, and information retrieval (Theme 1-1). Others, however, described AI as a potential partner, engaging with it directly and viewing it as a teaching assistant or a source of inspiration (Theme 3-2). This positioning is not fixed but dynamic, shaped by contextual resources and teachers' own orientations (Theme 3-3). In well-resourced settings with strong institutional support, teachers with an exploratory mindset were more inclined to treat AI as a co-creative partner (Theme 3-2); conversely, in resource-constrained contexts or where professional authority was emphasized, AI was used mainly in a utilitarian capacity (Theme 3-1). Overall, the “tool–partner” distinction operates less as a rigid dichotomy and more as a fluid continuum shaped by context (Theme 3-3).

Efficiency-driven AI use, with emerging innovation in interdisciplinary STEM field

Teachers' most frequent uses of AI remain efficiency-driven, such as drafting administrative documents (Theme 1-4), preparing test items, or compiling background information for science teaching (Theme 1-1). These practices reduce teachers' workload and allow more time for instructional guidance. At the same time, some teachers have begun to explore AI for more innovative applications (Theme 3-2). In science classes, AI was used to gather and compare disaster-related data to design inquiry-based tasks (Theme 1-2) that combined geography and engineering perspectives, such as analyzing resilient building designs across regions. In technology courses, AI supported students in programming or prototyping (Theme 1-2), helping them refine initial ideas into tangible project designs (Theme 2-1). At science fairs and in interdisciplinary projects, AI was employed to contextualize data (Themes 2-3) and generate visualizations (Theme 2-2), and even to create music or narrative scenarios for presentations (Theme 1-2), thereby enriching students’ multimodal communication. These practices show that AI in STEM education is gradually moving beyond a “time-saver” (Theme 3-1) toward a resource for cross-disciplinary integration and innovative learning (Theme 2-2).

Divergent patterns across STEM disciplines

Disciplinary cultures also shaped how teachers adopted and evaluated AI. Math and science teachers emphasized accuracy, noting that AI's high error rates in advanced problems (Theme 5-2) confined its role to generating items and diagrams rather than functioning as a dependable problem solver. By contrast, science teachers regarded AI as a valuable tool for integrating information (Theme 2-1) and expanding ideas (Theme 2-2), particularly in the design of science fair projects and inquiry-based activities (Theme 1-2). Teachers engaged in interdisciplinary projects went further, using AI to create multimodal materials (Theme 2-2) that enhanced student motivation and contextual engagement (Theme 4-1). Taken together, these contrasts highlight how disciplinary orientations influenced AI adoption: mathematics prioritized verification and precision (Theme 6-3), while natural sciences and interdisciplinary teaching placed greater value on integration and creativity (Theme 2-2).

Conclusions and Implications

This study reveals that generative AI adoption among Taiwanese STEM teachers is a dynamic process shaped by the interplay of individual attitudes, pedagogical practices, and disciplinary cultures. Teachers' perceptions fluctuate between viewing AI as a utilitarian tool and a co-creative partner, a transition largely dictated by environmental resources and specific instructional needs.

While current applications primarily focus on efficiency—such as administrative automation and material development—an evolutionary shift toward innovation is emerging. This is particularly evident in interdisciplinary contexts, where AI facilitates inquiry-based design, project prototyping, and multimodal content creation, evolving from a time-saving utility into a catalyst for deep learning. Crucially, disciplinary norms significantly influence adoption. Mathematics teachers, prioritizing precision, remain cautious due to potential errors. On the other hand, those in natural sciences and interdisciplinary fields leverage AI's capacity to integrate information and spark creativity, leading to more flexible instructional strategies.

To optimize AI's educational potential, professional development must abandon one-size-fits-all models. Instead, policy should provide differentiated, discipline-specific support that empowers educators to transcend mere tool users and embrace their potential as co-creative partners in the AI-integrated classroom.

References

Esiyok, E., Gokcearslan, S., & Kucukergin, K. G. (2024). Acceptance of Educational Use of AI Chatbots in the Context of Self-Directed Learning with Technology and ICT Self-Efficacy of Undergraduate Students. *International Journal of Human–Computer Interaction*, 1–10. https://doi.org/10.1080/10447318.2024.2303557

Kim, J., Lee, H., & Cho, Y. H. (2022). Learning design to support student-AI collaboration: Perspectives of leading teachers for AI in education. *Education and Information Technologies*, *27*(5), 6069–6104. https://doi.org/10.1007/s10639-021-10831-6

Luckin, R., Cukurova, M., Kent, C., & Du Boulay, B. (2022). Empowering educators to be AI-ready. *Computers and Education: Artificial Intelligence*, *3*, 100076. https://doi.org/10.1016/j.caeai.2022.100076

Miao, F., & Cukurova, M. (2024). *AI competency framework for teachers.* United Nations Educational, Scientific and Cultural Organization [UNESCO]. https://doi.org/10.54675/ZJTE2084

Nazaretsky, T., Ariely, M., Cukurova, M., & Alexandron, G. (2022). Teachers' trust in AI-powered educational technology and a professional development program to improve it. *British Journal of Educational Technology*, *53*(4), 914-931. https://doi.org/10.1111/bjet.13232

Popenici, S. A., & Kerr, S. (2017). Exploring the impact of artificial intelligence on teaching and learning in higher education. *Research and Practice in Technology Enhanced Learning*, *12*(1), 22. https://doi.org/10.1186/s41039-017-0062-8

Ramazanoglu, M., & Akın, T. (2024). AI readiness scale for teachers: Development and validation. *Education and Information Technologies*, 1-29. https://doi.org/10.1007/s10639-024-13087-y

Sanusi, I. T., Ayanwale, M. A., & Chiu, T. K. (2024). Investigating the moderating effects of social good and confidence on teachers' intention to prepare school students for artificial intelligence education. *Education and Information Technologies*, *29*(1), 273–295. https://doi.org/10.1007/s10639-023-12250-1

International Society of the Learning Sciences

The Relationship between Argumentative Discourse Sequences with Generative AI and Students' Argumentative Writing Structure

Kyungri Kim, Shingwan Elementary School, South Korea, kimkyungri11@gmail.com

Abstract: This poster examines how sixth-grade students' argumentative discourse with generative AI relates to the structure of their subsequent argumentative writing. Using Pragma-dialectical stages for discourse and Toulmin elements for essays, the study found that AI-mediated dialogue was dominated by confrontation-oriented claim and rebuttal sequences, while students' essays showed stronger transfer of data and backing than rebuttal, warrant, or qualifier use. The findings suggest that generative AI can support evidence-based argument writing when classroom prompts explicitly scaffold warrants, qualifiers, and reflection on opposing positions.

Key words: Generative AI, argumentative discourse, AI-learner interaction, sequence analysis, writing transfer

Introduction

Argumentation is central to critical thinking, yet elementary students often struggle to connect claims, evidence, warrants, and rebuttals in writing. Generative AI may function as a dialogic partner that prompts learners to state positions, respond to counterarguments, and elaborate evidence. This study asks: (1) What Pragma-dialectical sequence patterns appear in students' argumentative discourse with generative AI? (2) How are Toulmin elements distributed in their argumentative writing? (3) How are discourse sequence patterns related to writing structure?

Literature review

Pragma-dialectics Theory for Argumentative Discourse Analysis

Pragma-dialectical theory treats argumentation as an interactive process moving through confrontation, opening, argumentation, and concluding stages. This lens is useful for analyzing AI-mediated dialogue because it captures how learners initiate disagreement, negotiate standpoints, and respond to challenges.

Toulmin's Argumentative Model for Analyzing Writing Structure

Toulmin's model identifies claim, data, warrant, backing, qualifier, and rebuttal as core elements of argumentative structure. Prior writing research shows that students commonly produce claims and evidence but need support to articulate warrants, qualifications, and counterargument responses.

Generative AI in argumentative discourse and writing

Recent studies suggest that generative AI can promote higher-order thinking by asking students to justify, compare, and revise ideas. However, less is known about whether discourse patterns formed with AI transfer into students' independent argumentative writing, particularly in elementary classrooms.

Method

Twenty sixth-grade students in South Korea participated in six sessions using ChatGPT to discuss controversial issues. Students acted as discussants, while ChatGPT provided prompts, counterarguments, and follow-up questions. This poster reports preliminary analysis of four students' six-session logs and their argumentative essays.

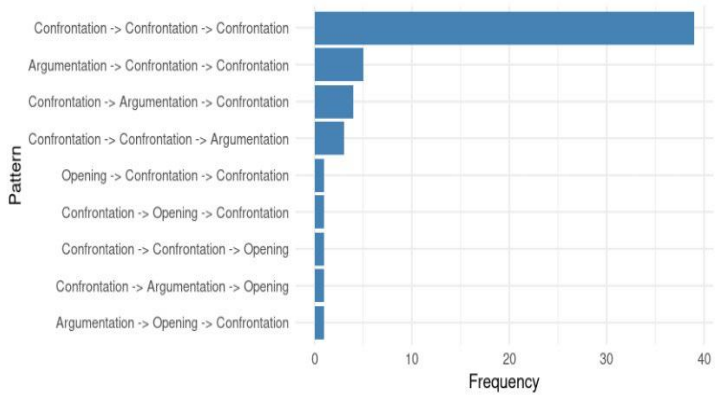
Discourse was coded at the utterance level using Pragma-dialectical stages, and essays were coded at the sentence level using Toulmin's six elements. Two trained coders independently coded the data and resolved disagreements through discussion.

Preliminary results

Across 95 utterances, discourse was dominated by Confrontation (84 instances, 88.4%), followed by Argumentation (9, 9.5%) and Opening (2, 2.1%). The most frequent 3-gram sequence was Confrontation-Confrontation-Confrontation, indicating repeated cycles of position-taking and rebuttal rather than progression through all stages of critical discussion.

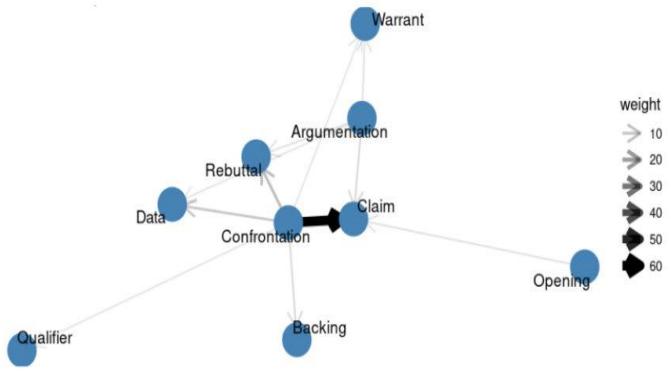
 © ISLS

Figure 1
Sequential Patterns of Discourse Stages (3-gram)



Toulmin coding of discourse showed Claim (66, 69%), Rebuttal (13, 14%), Data (9, 9%), Backing (4, 4%), Warrant (3, 3%), and Qualifier (1, 1%). In essays, Data (224, 41%) and Claim (114, 21%) were most frequent, followed by Backing (88, 16%), Warrant (57, 10%), Qualifier (52, 10%), and Rebuttal (12, 2%). Thus, AI dialogue appeared to support claim and evidence generation, but rebuttal and warranting transferred less consistently to writing.

Figure 2
Pragma-Dialectical Stages and Toulmin Elements Network



Discussion

The results suggest partial transfer from AI-mediated discourse to argumentative writing. Dialogue with generative AI supported repeated claim-rebuttal exchanges and helped students accumulate evidence, but students still needed explicit instruction to transform conversational opposition into written rebuttals, warrants, and qualifiers. For classroom use, AI argumentation tasks should require students to identify the strongest opposing view, explain why evidence supports a claim, and revise essays using warrants and qualifiers.

References

Borge, M., Smith, B. K., & Aldemir, T. (2024). Using generative ai as a simulation to support higher-order thinking. *International Journal of Computer-Supported Collaborative Learning, 19*(4), 479-532.

Cui, X., Ma, Z. Q., You, X. Y., Chen, Y., Yao, J. J., & Tu, Y. F. (2025). Incorporating generative conversational agents into collaborative argumentation: Do different customization strategies matter?.*Educational Technology & Society, 28*(2). 348-372.

Karbach, J. (1987). Using Toulmin's model of argumentation. *Journal of Teaching Writing, 6*(1), 81-92.

Larson, B. Z., Moser, C., Caza, A., Muehlfeld, K., & Colombo, L. A. (2024). Critical thinking in the age of generative AI. *Academy of Management Learning & Education, 23*(3), 373-378.

Liu, D. (2023). Pragma-dialectical perspective to intercultural discussion as communicative activity. *Humanities and Social Sciences Communications, 10*(1), 1-9.

 © ISLS

Magalhães, A. L. (2020). Teaching how to develop an argument using the Toulmin model.International *Journal of Multidisciplinary and Current Educational Research (IJMCER), 2*(3), 1-7.

OECD. (2019). *An OECD learning framework 2030. InThe future of education and labor(pp. 23-35)*. Cham: Springer International Publishing.

Tang, K. S. (2025). AI-textuality: Expanding intertextuality to theorize human-AI interaction with generative artificial intelligence.*Applied Linguistics*, amaf016.

Tang, K. S., & Putra, G. B. S. (2025). Generative AI as a Dialogic Partner: Enhancing Multiple Perspectives, Reasoning, and Argumentation in Science Education with Customized Chatbots. *Journal of Science Education and Technology*, 1-13.

Toulmin, S., Rieke, R. D., & Janik, A. (1979). An introduction to reasoning.

Van Eemeren, F. H., & Grootendorst, R. (1982). *Speech Acts in Argumentative Discussions: A theoretical model for the Analysis of Discussions directed towards solving conflicts of opinion.* Foris Publications.

Van Eemeren, F. H., & Grootendorst, R. (1995). The pragma-dialectical approach to fallacies.Fallacies: *Classical and contemporary readings*, 130-144.

Vincent-Lancrin, S., González-Sancho, C., Bouckaert, M., De Luca, F., Fernández-Barrerra, M., Jacotin, G., Urgel, J., & Vidal, Q. (2019). *Fostering Students' Creativity and Critical Thinking: What It Means in School. Educational Research and Innovation*. OECD Publishing. https://www.oecd.org/en/publications/fostering-students-creativity-and-critical-thinking_62212c37-en.html

 © ISLS

International Society of the Learning Sciences

Belonging Infrastructure: A Designable Construct for Cultivating Teacher Agency in RPPs

Gregory Benedis-Grab, University of Colorado Boulder, Greg.Benedis-Grab@colorado.edu
Alexandra Gendreau Chakarov, Colorado School of Mines, alexandra.chakarov@minese.edu

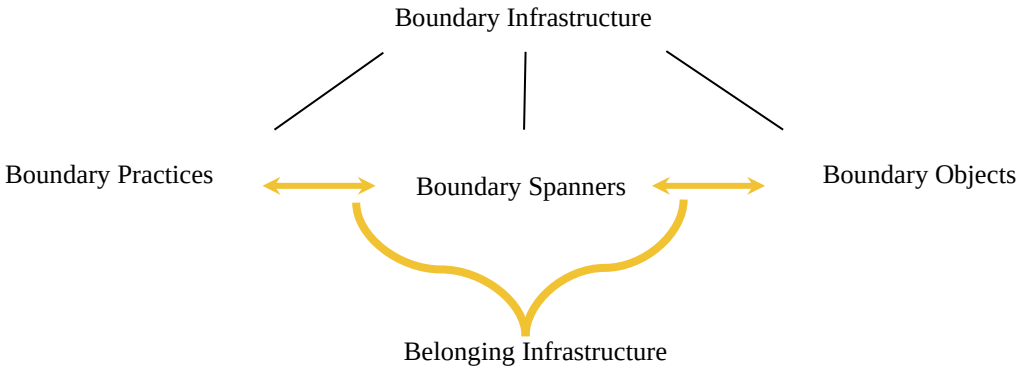
Abstract: Research-Practice Partnerships (RPPs) often rely on boundary infrastructure to explain how collaboration is organized, but this lens overlooks how teacher agency is cultivated. We introduce belonging infrastructure, the designable features of collaboration that legitimize participant contributions and make belonging visible. Evidence from a teacher co-design workshop demonstrates how this construct provides a new lens for intentionally designing partnerships that foster teacher leadership, build district capacity, and expand who feels authorized to teach computer science.

Introduction and theoretical framing

Research-Practice Partnerships (RPPs) are instrumental for improving education (Coburn & Penuel, 2016; Fishman et al., 2013), and the concept of boundary infrastructure, the routines, tools, and roles that mediate collaboration, helps explain how they function (Farrell et al., 2022). While a valuable framework, focusing on the structural elements, it offers less guidance on how to intentionally design for the relational dimensions within curricular co-design that enable teachers to move from being participants to integral partners (Gomez et al., 2022; Severance et al., 2018). This is a critical need when supporting educators teaching outside their primary expertise, a common challenge in elementary computer science (Ni et al., 2023; Yadav et al., 2016).

To address this design challenge, we introduce the construct of belonging infrastructure: the network of *designed participatory structures* (e.g., co-created norms, teacher-led artifacts, public critique routines) that validate participants' knowledge and legitimize their contributions (Figure 1). This infrastructure is what allows teachers to confidently engage with, critique, and create the boundary objects and practices of the partnership. This leads to our research question: *How does belonging infrastructure characterize the development of teacher boundary spanners in a research-practice partnership?*

Figure 1
Belonging Infrastructure Mediates the Engagement of Boundary Spanners with Boundary Practices and Objects



Methods

This study occurred in a new RPP among two local universities and a neighboring large, urban school district in the southwestern United States. We analyzed data from a four-day summer co-design workshop where four elementary STEM teachers and one district leader developed a third-grade physical computing unit using the micro:bit, a small, programmable microcontroller designed for introducing computing to young children (Sentance, et al., 2017). The workshop included intentionally designed participatory structures (e.g., co-created norms, teacher-authored artifacts, public critique). We analyze how belonging infrastructure emerged through these structures. Each day balanced design tasks with participatory structures intended to foster belonging: Day 1 built community through icebreakers, norms, and a co-created teacher goals board; Day 2 focused on shaping the

 © ISLS

storyline and guiding questions, with teachers reframing the unit design; Day 3 emphasized lesson development, live coding, and group reflection; Day 4 centered on design iteration and teacher-led critique of the unit structure.

Data sources included video recordings of the entire workshop and all physical artifacts created. We conducted a hybrid thematic analysis (Braun & Clarke, 2006), combining deductive and inductive approaches. Through the hybrid thematic analysis, we identified four observable dimensions of belonging infrastructure in practice: *belonging to the computer science community* (e.g., use of CS concepts, adaptations for learners, self-efficacy statements), *belonging within the RPP* (e.g., use of "we," expressions that ideas mattered, openness about uncertainty), *agency* (e.g., initiating or redirecting planning, proposing new goals or tools), and *navigating discomfort* (e.g., asking clarifying questions, reframing ideas after feedback). These categories guided our coding and interpretation of the workshop data.

Findings

These moments were enabled by designed structures that positioned teachers as contributors, including public artifact creation, flexible use of shared space, and opportunities to reorganize and critique the work. Our analysis shows how these designed participatory structures functioned as a belonging infrastructure. These designed structures created the conditions for teachers to take intellectual risks, demonstrate agency, and ultimately shape the partnership's work. We present two illustrative cases that show this construct in action through these structures (see Table 1). In the first case, a teacher's spontaneous drawing of a computer program became a new boundary object that advanced the group's collective understanding. In the second, a teacher demonstrated design agency by physically reorganizing the unit timeline, a move that reshaped a core boundary practice of the RPP. These moments, detailed in Table 1, illustrate how belonging was not just an internal feeling but was enacted and made visible through recognized contributions. The belonging infrastructure provided the means for teachers to author new roles and become integral partners in the co-design process.

Table 1

Illustrative Cases of the Belonging Infrastructure in Action

Case 1: Creating a Boundary Object	Case 2: Demonstrating Design Agency
During a session focused on the "sound-o-meter" program, one teacher spontaneously created a visual model to clarify its function. Her drawing included branching arrows for conditional logic and a looping arrow for a counting variable, representing the complex program in a simple, accessible way. At the time, the group had been grappling with how to represent the program's logic for students. This teacher's model served not only to clarify the program for herself but for the entire group. It became a shared reference point, a new boundary object, for subsequent discussions, allowing the team to develop a more concrete, collective understanding. The drawing was posted on the wall and later praised by a district leader for its clarity and potential as a student-facing tool.	During a planning session, the group discussed how to structure the unit's storyline. As the conversation unfolded, one teacher stood up, took several sheets of chart paper, and began creating a visual timeline on the wall. On each page, she wrote the guiding question for each lesson, spatially organizing the unit to clarify its narrative arc. This simple but powerful move demonstrated design agency. It transformed an abstract conversation into a concrete planning scaffold that the group immediately adopted. Her action reshaped the boundary practice of lesson planning and solidified her role not just as a participant, but as a co-designer who could structure and lead the group's work.

Discussion and significance

This study introduces belonging infrastructure to characterize the development of teacher boundary spanners as increasing participation in shaping boundary objects and practices. It does so by illuminating the moment-to-moment interactions where teachers, through acts of design agency, create new boundary objects and reshape boundary practices. This construct extends the theory of boundary infrastructure by shifting focus from the mere existence of collaborative structures to the designable conditions that invite all participants to become integral partners. Unlike boundary infrastructure, which describes coordination structures, belonging infrastructure focuses on *designing for visible participation and authorship*. Ultimately, belonging infrastructure offers a critical lens for designing more equitable partnerships that cultivate teacher leadership by recognizing them as active contributors to system change and knowledge building. We invite discussion on how belonging infrastructure may surface in other RPP contexts, and whether making it visible helps expand who is recognized as someone whose ideas and expertise matter in education partnerships

 © ISLS

References

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative research in psychology, 3(2),77-101.

Coburn, C. E., & Penuel, W. R. (2016). Research–practice partnerships in education: Outcomes, dynamics, and open questions. Educational researcher, 45(1), 48-54.

Farrell, C. C., Penuel, W. R., Allen, A., Anderson, E. R., Bohannon, A. X., Coburn, C. E., & Brown, S. L. (2022). Learning at the boundaries of research and practice: A framework for understanding research–practice partnerships. Educational Researcher, 51(3), 197-208.

Fishman, B. J., Penuel, W. R., Allen, A. R., Cheng, B. H., & Sabelli, N. O. R. A. (2013). Design-based implementation research: An emerging model for transforming the relationship of research and practice. Teachers College Record, 115(14), 136-156.

Gomez, K., Lee, U. S., & Woodman, A. B. (2022). The role of teacher beliefs, goals, knowledge, and practices in co-designing computer science education curricula. In Teacher Learning in Changing Contexts (pp. 158-174). Routledge

Ni, L., Bausch, G., & Benjamin, R. (2023). Computer science teacher professional development and professional learning communities: A review of the research literature. Computer Science Education, 33(1), 29-60.

Sentance, S., Waite, J., Yeomans, L., & MacLeod, E. (2017, November). Teaching with physical computing devices: the BBC micro: bit initiative. In *Proceedings of the 12th workshop on primary and secondary computing education* (pp. 87-96).

Severance, S., Penuel, W. R., Sumner, T., & Leary, H. (2018). Organizing for teacher agency in curricular co design. In Cultural-historical activity theory approaches to design-based research (pp. 45-78). Routledge.

Yadav, A., Gretter, S., Hambrusch, S., & Sands, P. (2016). Expanding computer science education in schools: understanding teacher experiences and challenges. Computer science education, 26(4), 235-254.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grants No. 2434834 and 2434835. The opinions expressed are those of the authors and do not represent views of the NSF.

International Society of the Learning Sciences

Resolving Conceptual Contradictions through Transformative Integration

Weiyi Xiao, Independent Researcher, xiaoweiyi25@gmail.com

Abstract: This paper proposes transformative integration as a more desirable approach to resolving contradictions in contemporary educational collaboration. It conceptualizes transformative integration as the generation of thirdness and operationalizes thirdness as a Principle of Transformation (PoT) or as an actionable solution derived from it. In doing so, the paper lays the groundwork for systematically identifying and intentionally cultivating thirdness in practice, thereby supporting creativity-oriented collaboration and students' meaningful participation in democratic, innovation-driven societies.

Introduction

Students learn collaboratively by working through differences. Deeper learning occurs when they confront conceptual contradictions expressed as direct opposition or divergent perspectives. In responding to such tensions, argumentative and summative integration are often prioritized over defense due to their collaborative orientation. While argumentative integration resolves disagreement by specifying the contexts in which particular claims hold (Leitão, 2000), summative integration abstracts a higher-level synthesis from overlapping points across conflicts (Scardamalia, 2002).

Yet these approaches seem suboptimal in contemporary educational collaboration. Due to their emphasis on reasoning within given frameworks or generalizing from shared elements, they produce reactive insights tied to immediate or similar contexts (Engeström, 2014). As situations evolve and unfamiliar challenges emerge, however, these insights may require continual revision to stay relevant, which can become intellectually exhausting in an increasingly diverse and rapidly evolving world, prompting disengagement or even defensiveness rather than openness to novel experiences in practice.

To address this limitation, collaborative learning should move toward transformative integration that cultivates students' capacity to anticipate and adapt to change proactively. This paper introduces thirdness to describe such an approach. Drawing on relevant literature, it conceptualizes and operationalizes thirdness so that this construct can be more systematically identified and deliberately fostered in collaborative practice. In doing so, the paper seeks to promote creativity-oriented collaboration more intentionally in education and support students' more meaningful participation in democratic, innovation-driven societies.

Conceptualizing thirdness in education

Thirdness is a term introduced by Peirce in phenomenology, where it represents the law or principle that mediates Firstness (i.e., the pure, immediate quality of experience) and Secondness (i.e., reaction to that immediacy) (Sonesson, 2013). In education, the term appears in a non-capitalized form in expansive learning theory, where it denotes a "germ cell" concept in the form of an abstract principle that provides a basic explanatory mechanism for the origin of the contradiction under investigation and can be expanded into multiple concrete manifestations that push the initial problematic situation into new phases of development (Engeström, 2014).

In collaborative learning, thirdness may be explored through the notion of principled practical knowledge (PPK) (Bereiter, 2009, 2014), a type of internally coherent, generalizable, and systematic knowledge produced to solve problems creatively in knowledge-poor situations and capable of deducting novel conclusions that help a field of practice advance. It has also been discussed in dialogic education, where it refers to a position that constitutes a dialogic space in which different ideas converge in some respects while diverging in others, sparking the emergence of new meanings that allow people to view a situation from at least two standpoints at once and to expand the context to look at viewpoints beyond the immediate community (Wegerif, 2013, 2018).

These discussions indicate an early recognition and conceptualization of thirdness as a powerful means of resolving contradictions in inclusively integrative, collectively transformative, and expansively informative ways, helping students generate new understandings that enable them to adapt to novel practices proactively.

Operationalizing thirdness in education

Nevertheless, although earlier discussions provide concrete examples of thirdness, these instances remain disconnected and largely descriptive, lacking identification of the underlying patterns that would allow the

 © ISLS

construct to be analytically recognized or intentionally cultivated. To become a more useful construct in collaborative learning practice, thirdness must be operationalized by specifying its structural features.

Since thirdness is an abstract principle that both addresses and transcends existing contradictions, the extended abstract principle described in SOLO (the Structure of the Observed Learning Outcome) taxonomy (Biggs & Collis, 1982) may provide a useful framework. When applied to reconcile conflicting perspectives, the principle situates the ideas within a relativizing, concept-based abstract structure that shapes and sustains open-ended coherence within and beyond the given context. It interprets the immediate, divergent cases as contextual variations of this structure and enables further deduction from this structure to generate new contextual variations. This paper refers to this structure as a Principle of Transformation (PoT), which is often expressed in statements such as "system 1 varies with system 2" or "variations in system 1 depend on variations in system 2."

Nevertheless, since a PoT holds conclusions open, it remains limited as a form of theoretical resolution and may not constitute a genuine resolution in contexts where contradictions are settled only through actionable solutions (e.g., in an engineering class). To make thirdness meaningful across learning situations, this paper extends prior work by operationalizing thirdness also as an actionable solution derived from the PoT. When a collaborative setting does not require a practical outcome, thirdness is achieved through the proposal of a PoT. When a practical outcome is required, a workable solution derived from the PoT completes the creative process. Table 1 illustrates the different types of integration through examples, while Figure 1 uses vector-like diagrams to highlight their differences, with arrows representing conceptual directions within a dialogic space.

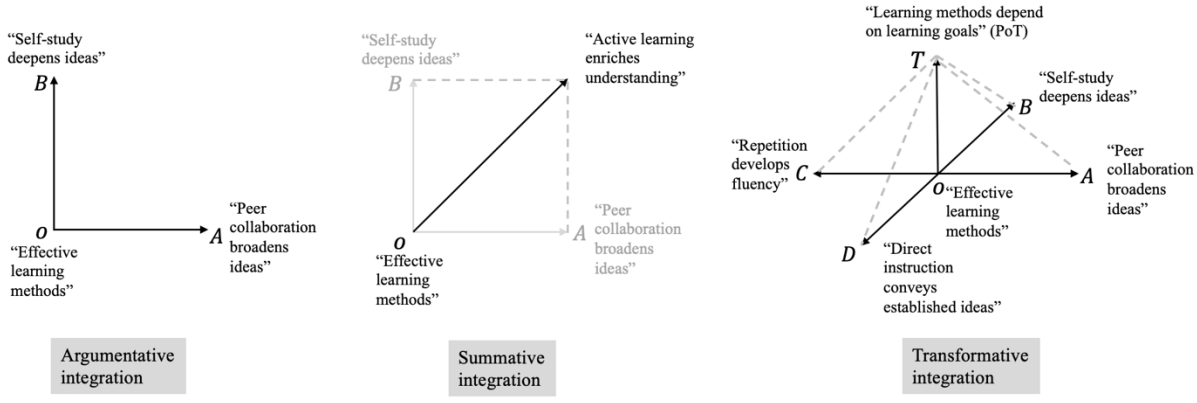
Table 1

Argumentative, Summative, and Transformative Integration with Simplified Reasoning

Initial contradiction	Which learning approach is more effective: self-study or peer interaction?
Argumentative integration	Self-study deepens understanding, whereas peer interaction broadens perspectives.
Summative integration	Learning is effective when learners engage actively with the material.
Transformative integration-theoretical thirdness	The effectiveness of a learning approach depends on the goal: self-study deepens understanding, peer interaction broadens perspectives, direct instruction conveys established knowledge, and repeated practice develops procedural fluency.
Transformative integration-actionable thirdness	The effectiveness of a learning approach depends on the goal … (same as above) … To be effective, we need to adapt our approach as our learning goals shift.

Figure 1

Illustrations of Different Types of Integration



Conclusion

To make collaborative learning more meaningful in today's constantly changing world, this paper introduces thirdness as a more desirable approach to addressing conceptual contradictions. Although such conceptual work represents only a starting point, it provides an important foundation for the intentional cultivation of creativity-oriented collaboration in education and for students' meaningful participation in contemporary democratic, innovation-driven societies.

 © ISLS

References

Bereiter, C. (2009). Innovation in the absence of principled knowledge: The case of the Wright Brothers. *Creativity and Innovation Management, 18*(3), 234–241. https://doi.org/10.1111/j.1467-8691.2009.00528.x

Bereiter, C. (2014). Principled practical knowledge: Not a bridge but a ladder. *Journal of the Learning Sciences, 23*(1), 4–17. https://doi.org/10.1080/10508406.2013.812533

Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning: The SOLO taxonomy (Structure of the observed learning outcome)*. Academic Press.

Engeström, Y. (2014). *Learning by expanding: An activity-theoretical approach to developmental research* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781139814744

Leitão, S. (2000). The potential of argument in knowledge building. *Human Development, 43*(6), 332–360. https://doi.org/10.1159/000022695

Scardamalia, M. (2002). Collective cognitive responsibility for the advancement of knowledge. In B. Smith (Ed.), *Liberal education in a knowledge society* (pp. 67–98). Open Court.

Sonesson, G. (2013). The natural history of branching: Approaches to the phenomenology of Firstness, Secondness, and Thirdness. *Signs and Society, 1*(2), 297–325. https://doi.org/10.1086/673251

Wegerif, R. (2013). *Dialogic: Education for the Internet age*. Routledge. https://doi.org/10.4324/9780203111222

Wegerif, R. (2018). A dialogic theory of teaching thinking. In K. Laura & R. Wegerif, *Theory of teaching thinking* (1st ed., pp. 89–104). Routledge. https://doi.org/10.4324/9781315098944-7

 © ISLS

International Society of the Learning Sciences

Rotating Leadership of Transactive Discourse through Synchrony and Desynchrony of Multimodal Interaction

Jun Lu, Jun Oshima, Ritsuko Oshima
ro.takashi.18@shizuoka.ac.jp, joshima@inf.shizuoka.ac.jp, roshima@inf.shizuoka.ac.jp
Graduate School of Science and Technology, Shizuoka University, Japan

Abstract: *Rotating* leadership emerges in successful collaboration. This study investigates the extent to which their multimodal interactions are synchronized or not, along with the *rotating* leadership of transactive discourse (LoT). Nine groups of three members participated in a 45-minute co-located CPS task, and their spatial positions were collected. The results suggest that in addition to the synchronization of the spatial positioning, *desynchronization* was observed in *rotating* LoT. The potential use of the kinesthetic modality is discussed.

Introduction

Rotating leadership can be observed in successful collaborative activity, as one form of shared leadership. Successful collaboration is supported by shared agency. Thus, the leadership is emergent, and the temporal leader is not *fixed* but *rotates* (Ma et al., 2016). In this study, we investigate *rotating* leadership from the perspective of transactivity (Teasley, 1997). Transactivity is a social mode of knowledge co-construction and has been considered vital for productive collaborative problem-solving (CPS) (Vogel et al., 2023; Weinberger & Fischer, 2006). Previous studies have operationalized the leadership of transactive discourse (LoT) as an engagement to the discourse moves in which ideas are connected to the existing discussions (Oshima et al., 2023; Lu et al., 2025). While transactivity and *rotating* LoT are considered one of the important aspects to understand productive CPS, previous studies suggest that multimodal synchrony may also influence the CPS productivity and the transactive discourse (Li et al., 2022; Aoyama Lawrence & Weinberger, 2022; Nasir et al., 2021; Popov et al., 2017). In this study, we shed light on multimodal interactions in *rotating* LoT.

To capture and analyze multimodal interactions in co-located collaboration, it is important to define and integrate the modalities appropriately (Esterhazy et al., 2025; Schneider, 2023). Esterhazy et al. (2025) proposed the Multimodal Collaboration Analytics (MMCA) framework and defined five categories of modality; here, we address the auditory, visual, and kinesthetic modalities, excluding the tactile and physiological ones, as our interest lies in dynamic group interaction.

Previous studies suggested that the synchrony of auditory and visual modalities is potentially crucial for transactive discourse (Nasir et al., 2021; Aoyama Lawrence & Weinberger, 2022; Sharma & Molinari, 2023). However, the kinesthetic modality remains underexplored. Kinesthetic modality, such as spatial positioning, may shed light on embodied teamwork in co-located collaboration (Zhao et al., 2023). In contrast, Popov et al. (2017) argued that synchrony of action might be important but not significant in transactive discourse. Further investigation is required on the synchrony of the kinesthetic modality to uncover the multimodal interplay in the transactive discourse. In this study, we employ spatial position as a kinesthetic modality to analyze the synchrony in *rotating* LoT. Specifically, we address the following research question: To what extent does the synchrony of spatial positioning differ between *rotating* and *fixed* LoT?

Methods

Nine groups of three undergraduates, 27 in total, at a Japanese university participated in a CPS task. The task was from the Jasper series (CTGV, 1992), and participants were asked to watch a 17-minute video storytelling before the 45-minute CPS session. During the CPS session, they had access to two artifacts: a computer and a whiteboard. The computer was available for watching the video, and the whiteboard was available for writing and organizing information. The objective of the task was to consider possible routes to bring the eagle to an animal hospital located in town as fast as possible. The whole CPS session was recorded and transcribed. A stereo camera was also installed for spatial position analysis.

We analyzed the temporal trajectory of the LoT based on the transcripts. Following previous studies (Lu et al., 2025; Oshima et al., 2023), the group's transactive discourse was modeled as a word co-occurrence network, and the individual LoT is operationalized as the contribution to the network. We extracted task-related words from the transcripts as the nodes in the network. These words (nodes) are linked through the conversation turns by transactive discourse moves, such as integration and conflict resolution, which can be operationalized as the co-occurrence of words. In addition, we operationalized the individual LoT based on Lu et al. (2025), in which LoT

 © ISLS

is defined as the extent to which the nodes and links added by an individual contributed to the development of the network. The temporal trajectory was captured by the conversation turns, using KBDeX (Oshima et al., 2012).

To extract spatial position information, we developed a stereo capture system and conducted 3D skeleton detection. A stereo camera was placed in front of the artifacts, capturing 3D images of all three members. Their spatial position was automatically annotated with three labels: in front of the computer (*computer*), in front of the whiteboard (*whiteboard_front*), and around the whiteboard (*whiteboard_around*).

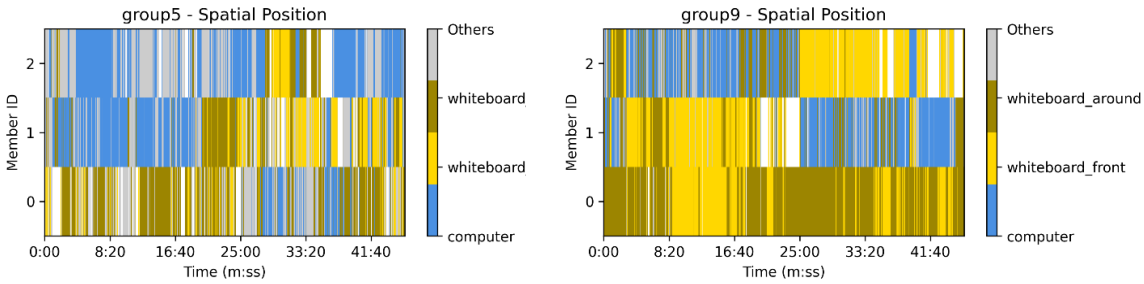
Results and discussion

The results suggested that the dyadic synchrony was frequently reconfigured in the groups with *rotating* LoT, while there were fewer reconfigurations in the groups with *fixed* LoT. The temporal trajectories of LoT suggested that *rotating* LoT emerged in groups 1, 2, 3, and 5, implying collective contribution to the group's knowledge co-construction. Conversely, groups 4, 6, and 8 were led by two of the three members, and groups 7 and 9 were led by only one member, both implying unbalanced contributions. In the following sections, we compare the members' spatial positioning between groups 5 and 9, where significant *rotating* and *fixed* LoT were observed, respectively.

Figure 1 shows the distribution of spatial position among members in groups 5 and 9, suggesting the different patterns of dyadic synchrony. Figure 1(a) suggests frequent reconfiguration of dyadic synchrony in group 5, where *rotating* LoT is demonstrated. Dyadic synchrony was first observed between members 1 and 2 around the computer, then member 2 reconfigured the synchrony with member 0 around the whiteboard. After a while, member 0 moved in front of the computer and was in sync with member 2. Several reconfigurations were then observed among all members. The frequent reconfiguration suggests that members' engagement was actively self-organized and emerged. In contrast, group 9, where a *fixed* LoT was demonstrated by member 2, experienced less reconfiguration of dyadic synchrony, as illustrated in Figure 2(b). While members 1 and 2 both reconfigured their dyadic synchrony by switching their spatial position, member 0 stuck to the whiteboard and never interacted with the computer.

Figure 1

Distribution of Spatial Position for Each Member in Groups 5 (a) and 9(b)



(a) *Rotating* LoT among Three Members (b) *Fixed* LoT by Member 2

Note. Each row corresponds to each member, and colored bars represent the automatically-annotated spatial position (blue = in front of the computer, yellow = in front of the whiteboard, gold = around the whiteboard).

We argue that *desynchronization* of dyadic synchrony would be influential to the collaboration quality. Here, we define *desynchronization* as leaving synchrony and interacting with another member(s). Collaboration requires not only synchronous work but also asynchronous division of labor (Dillenbourg, 1999), and members' actions are sometimes distributed (e.g., Popov et al., 2017), based on their shared agency (Damșa, 2010). From the perspective of embodied teamwork (Zhao et al., 2023), members' shared agency is embodied, which could lead to distributed multimodal interactions. In terms of MMCA, it would be said that as a result of shared agency, *desynchronization* occurs in the kinesthetic modality, and *rotation* of LoT occurs in the verbal modality. Therefore, *desynchronization* might be another indicator of collaboration, along with the *rotating* LoT.

Our study contributes to the transactivity research by adding empirical implications that collective engagement to transactive discourse could be embodied not only in verbal and visual modalities but also in the kinesthetic modality. In addition, our methodological contributions also allow us to quantify the multimodal interaction in transactive discourse, implying the potential importance of *desynchrony*.

Despite the implications, several limitations remain. Due to the small sample size and coarse temporal granularity, the causal relationship is unclarified. In addition, the kinesthetic modality includes actions and body movements, which were not analyzed. Further research is required to explore *desynchronization* in more detail.

 © ISLS

References

Aoyama Lawrence, L., & Weinberger, A. (2022). Being in-sync: A multimodal framework on the emotional and cognitive synchronization of collaborative learners. *Frontiers in Education*, *7*, 867186.

Cognition and Technology Group at Vanderbilt. (1992). The Jasper series as an example of anchored instruction: Theory, program description, and assessment data. *Educational Psychologist*, *27*(3), 291–315.

Damşa, C., Kirschner, P. A., Andriessen, J. E. B., Erkens, G., & Sins, P. H. M. (2010). Shared epistemic agency: An empirical study of an emergent construct. *Journal of the Learning Sciences*, *19*(2), 143–186.

Dillenbourg, P. (1999). What do you mean by collaborative learning? In P. Dillenbourg (Ed.), *Collaborative learning: Cognitive and Computational Approaches* (pp. 1–19). Elsevier.

Esterhazy, R., Kaliisa, R., Sanchez, D., Langford, M., & Damşa, C. (2025). Advancing Multimodal Collaboration Analytics: A Scoping Review. *Journal of Learning Analytics*, 1–20.

Li, S., Pöysä-Tarhonen, J., & Häkkinen, P. (2022). Patterns of action transitions in online collaborative problem solving: A network analysis approach. *International Journal of Computer-Supported Collaborative Learning*, *17*(2), 191–223.

Lu, J., Oshima, J., & Oshima, R. (2025). Interrelation Between Leadership of Transactivity and "Equal" Participation in Collaborative Problem-Solving. In Oshima, J., Chen, B., Vogel, F., & Järvelä, J. (Eds.), Proceedings of the 18th International Conference on Computer-Supported Collaborative Learning - CSCL 2025 (pp. 678-680). International Society of the Learning Sciences.

Ma, L., Matsuzawa, Y., & Scardamalia, M. (2016). Rotating leadership and collective responsibility in a grade 4 Knowledge Building classroom. *International Journal of Organisational Design and Engineering*, *4*(1/2), 54.

Nasir, J., Kothiyal, A., Bruno, B., & Dillenbourg, P. (2021). Many are the ways to learn identifying multi-modal behavioral profiles of collaborative learning in constructivist activities. *International Journal of Computer-Supported Collaborative Learning*, *16*(4), 485–523.

Oshima, J., Oshima, R., & Matsuzawa, Y. (2012). Knowledge Building Discourse Explorer: A social network analysis application for knowledge building discourse. *Educational Technology Research and Development*, *60*(5), 903–921.

Oshima, J., Oshima, R., Yamashita, S., & Lu, J. (2023). Leadership of transactive discourse in collaborative problem solving. In Damşa, C., Borge, M., Koh, E., & Worsley, M. (Eds.), *Proceedings of the 16th International Conference on Computer-Supported Collaborative Learning - CSCL 2023* (pp. 360–361). International Society of the Learning Sciences.

Popov, V., van Leeuwen, A., & Buis, S. C. A. (2017). Are you with me or not? Temporal synchronicity and transactivity during CSCL. *Journal of Computer Assisted Learning*, *33*(5), 424–442.

Schneider, B. (2023). Three challenges in Implementing multimodal learning analytics in real-world learning environments. *Learning: Research and Practice*, *10*(1), 103–112.

Sharma, K., & Molinari, G. (2023). What Happens When Collaborators Are Not In-Synch?. In *Proceedings of the 16th International Conference on Computer-Supported Collaborative Learning-CSCL 2023*, pp. 99-106. International Society of the Learning Sciences.

Teasley, S.D. (1997). Talking About Reasoning: How Important Is the Peer in Peer Collaboration?. In *Springer eBooks* (pp. 361–384). https://doi.org/10.1007/978-3-662-03362-3_16

Vogel, F., Weinberger, A., … & Noroozi, O. (2023). Transactivity and knowledge co-construction in collaborative problem solving. In Damşa, C., Borge, M., Koh, E., & Worsley, M. (Eds.), *Proceedings of the 16th International Conference on Computer-Supported Collaborative Learning - CSCL 2023* (pp. 337-346). International Society of the Learning Sciences.

Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & Education*, *46*(1), 71–95.

Zhao, L., Swiecki, Z., Gasevic, D., Yan, L., Dix, S., Jaggard, H., Wotherspoon, R., Osborne, A., Li, X., Alfredo, R., & Martinez-Maldonado, R. (2023). METS: Multimodal Learning Analytics of Embodied Teamwork Learning. *ACM International Conference Proceeding Series*, 186–196.

Acknowledgement

JSPS KAKENHI Grant Number 23H01000 and JST SPRING Grant Number JPMJSP2167 supported this work.

International Society of the Learning Sciences

Dialogic Intelligence: Rethinking the pedagogical goal of CSCL in the context of Generative AI

Rupert Wegerif, Xuanning Chen, Bo Yu, Imogen Casebourne, Kevin Martin
Rw583@cam.ac.uk; xc330@cam.ac.uk, by283@cam.ac.uk, ic407@hughes.cam.ac.uk, kcm38@hughes.cam.ac.uk
University of Cambridge

Abstract: As generative AI surpasses students in knowledge delivery and examination performance, education faces an unprecedented challenge in defining what human intelligence means. This theory paper argues that our response should be to reclaim, revalue, and rethink intelligence itself, advancing dialogic intelligence as a central educational aim. The dominant view of intelligence as an individual, measurable trait reflects print literacy's historical affordances, which have obscured intelligence's fundamentally relational nature. AI's emergence as a ubiquitous cognitive technology necessitates recovering an older understanding: intelligence is the capacity to support the emergence of truth and goodness through responsive relationship—whether between people, with nature, or with technology such as generative AI. While the CSCL community has explored teaching thinking, research has typically focused on collaboration as a means to enhance curricular learning. The dialogic responsiveness of generative AI suggests instead that we should design CSCL to promote dialogic intelligence as an end in itself. We illustrate this agenda through two CSCL design examples that prioritize dialogic intelligence over content mastery.

Generative AI and educational aims

Generative AI has forced education into a moment of reckoning. When systems such as ChatGPT can pass professional examinations, retrieve information instantly, generate plausible arguments, and simulate dialogue, it becomes difficult to defend educational models centred on individual recall, routine problem-solving, and the production of competent prose. Much early response has been defensive: detecting AI use, designing "AI-proof" assessments, or treating AI simply as a support tool. But these responses avoid the deeper question for the Computer-Supported Collaborative Learning (CSCL) community: what should education aim to cultivate in human learners?

CSCL should move beyond treating collaboration mainly as a means of helping individuals master curriculum content and instead place dialogic intelligence at the centre of educational design. Dialogic intelligence is not an individual trait measured by speed, accuracy, or test performance. It is the capacity for meaning, insight, and judgement to emerge through responsive relationships: between people, between people and traditions of knowledge, and increasingly between people and AI systems.

Historically, dominant ideas of intelligence were shaped by print culture. Earlier traditions linked intelligence to judgement, discernment, and wisdom exercised in context. Socrates, for example, valued the "living word" of dialogue because it could respond, be questioned, and remain open to revision, unlike the fixed and answerless word of writing (Plato, *Phaedrus*, 275d). Over time, however, writing and later print culture encouraged a more monologic conception of thought: linear, abstract, decontextualized, and increasingly located within the individual mind (Ong, 1982; Olson, 1994; Toulmin, 1990). By the twentieth century, psychometric testing turned this into a scientific orthodoxy. Intelligence became whatever could be standardized, measured, and compared across individuals (Binet & Simon, 1916/1973; Spearman, 1904).

Yet this was never a neutral description of human thinking. It reflected the cognitive demands of print-based schooling. Tasks involving vocabulary, abstract classification, analogy, and formal reasoning were elevated as signs of intelligence because they fitted the dominant educational system. Cross-cultural studies and the Flynn effect later showed how contingent this was: what counts as intelligence changes with cultural practices, technologies, and forms of education (Ceci, 1991; Cole & Scribner, 1974). IQ captures something real, but that "something" is not intelligence in the broadest human sense. It is performance within a historically specific literacy regime.

Much CSCL research has significantly advanced understanding of how learners work together productively. However, it has largely remained tied to the print paradigm, treating collaboration as valuable because it improves individual outcomes such as knowledge acquisition, problem-solving, or reasoning performance. The deeper possibility is that intelligence does not merely happen inside individuals and then get expressed in groups. Rather, intelligence can emerge in the shared space of dialogue itself.

 © ISLS

Research on collective intelligence points in this direction. Woolley and colleagues showed that groups can display a form of collective intelligence not reducible to the abilities of individual members (Woolley et al., 2010). Work in the *Thinking Together* tradition similarly found that when students were taught norms for exploratory talk, both group and individual reasoning improved (Mercer & Littleton, 2007). Close analysis suggested that explicit reasoning moves alone did not explain the most important moments of insight. Often the decisive shift came in shared attention, pause, tension, or a sudden moment of recognition before explanation could even be articulated. The talk moves mattered, but mainly because they helped open what came to be called dialogic space: the relational field in which new meaning becomes possible (Wegerif, 2008).

Dialogic intelligence as an educational aim

Dialogic intelligence is not best understood as a possession but as an event. It arises when participants hold perspectives in tension, attend responsively to one another, and remain open to something beyond their current assumptions. Bakhtin's account of double-voicedness helps here: to think dialogically, one must inhabit one's own perspective while also entering into another's (Bakhtin, 1981). Mead's notion of the generalized other adds that reflective thought depends on taking up the standpoint of a wider community (Mead, 1934/1967). But genuine dialogue goes further than simply reproducing communal norms. It also opens toward what Bakhtin called a superaddressee and what might be described more broadly as an Infinite Other: a horizon of truth, goodness, or understanding that is never fully exhausted by current knowledge (Levinas, 1969; Wegerif, 2025).

Generative AI does not possess lived subjectivity, but it can function as a remarkably articulate simulated voice. It can represent the accumulated discourse of disciplines, professions, and traditions. In that sense, AI can give learners access to a powerful form of the generalized other. A student in dialogue with AI can encounter what counts as established understanding in mathematics, science, history, or law. The educational value of this does not lie in receiving better answers more efficiently. It lies in using that encounter as a platform for further questioning. If students simply defer to AI, they risk conformity to cliché. If they engage it dialogically, they can use it to test assumptions, reframe problems, and move beyond what is already established.

This distinction can be illustrated by the provisional findings of two design-based research examples which we have conducted at Cambridge. In one study, AI moderators were introduced into small-group collaborative learning. Early versions failed because they addressed students individually and did not support genuinely shared inquiry. After iterative refinement, however, students reported that the AI began to connect viewpoints, bring ideas together, and raise possibilities the group had not considered. It appeared to support depth, breadth, momentum, and inclusion within the discussion. Importantly, these gains did not register on standard critical thinking measures, suggesting that dialogic intelligence may require different forms of assessment from those used for decontextualized reasoning (Michaels & O'Connor, 2015; Hennessy et al., 2016).

In a second project, AI was designed as a dialogic partner for leadership training in difficult workplace conversations. Rather than acting as a compliant assistant, the system was configured to challenge, resist, and force participants to reconsider their habitual responses. Participants reported that this kind of difficult AI interlocutor was valuable precisely because it created the tension needed for growth. In both studies, the key issue was not AI capability in itself, but the quality of the relational space created through interaction.

These examples also clarify the difference between dialogic intelligence and what is often called hybrid intelligence. Hybrid intelligence usually refers to combining the strengths of humans and AI so that each compensates for the other's weaknesses. That is useful, but dialogic intelligence points to something deeper. The most important educational moments do not occur because two separate agents contribute separate capacities. They occur when a shared space of inquiry opens in which new understanding emerges that belongs fully to neither the human nor the machine alone.

The implication for CSCL is substantial. The task is no longer simply to design collaboration so that learners achieve predefined curricular goals more efficiently. It is to design environments that open, sustain, and deepen dialogic space. That means framing AI not mainly as answer-provider or tutor, but as a dialogic partner that can represent established knowledge while also provoking learners to move beyond it. It also means rethinking assessment so that it attends to relational capacities: how learners connect perspectives, sustain uncertainty, reshape questions, and generate insight together.

Generative AI exposes the narrowness of the print-based conception of intelligence on which much modern education has relied. What remains distinctly human is not superior information processing, but care: the capacity to care about truth, to remain answerable for judgement, and to sustain inquiry beyond what is immediately given. If CSCL can make dialogic intelligence its central aim, AI becomes not a threat to education but an opportunity to recover and extend intelligence in its deeper sense: as participation in a shared, open-ended dialogue through which we learn, create, and orient ourselves towards what is true and good.

References

Bakhtin, M. M. (1981). *The dialogic imagination: Four essays* (C. Emerson & M. Holquist, Eds.; M. Holquist, Trans.). University of Texas Press.

Binet, A., & Simon, T. (1973). *The development of intelligence in children* (E. S. Kite, Trans.). Arno Press. (Original work published 1916).

Ceci, S. J. (1991). How much does schooling influence general intelligence and its cognitive components? A reassessment of the evidence. *Developmental Psychology, 27*(5), 703–722.

Cole, M., & Scribner, S. (1974). *Culture and thought: A psychological introduction*. Wiley.

Hennessy, S., Rojas-Drummond, S., Higham, R., Márquez, A. M., Maine, F., Ríos, R. M., García-Carrión, R., Torreblanca, O., & Barrera, M. J. (2016). Developing a coding scheme for analysing classroom dialogue across educational contexts. *Learning, Culture and Social Interaction, 9*, 16–44. https://doi.org/10.1016/j.lcsi.2015.12.001

Levinas, E. (1969). *Totality and infinity: An essay on exteriority* (A. Lingis, Trans.). Duquesne University Press.

Mead, G. H. (1967). *Mind, self, and society from the standpoint of a social behaviorist* (C. W. Morris, Ed.). University of Chicago Press. (Original work published 1934).

Mercer, N., & Littleton, K. (2007). *Dialogue and the development of children's thinking: A sociocultural approach*. Routledge.

Michaels, S., & O'Connor, C. (2015). Conceptualizing talk moves as tools: Professional development approaches for academically productive discussions. In L. B. Resnick, C. Asterhan, & S. N. Clarke (Eds.), *Socializing intelligence through talk and dialogue* (pp. 333–347). AERA. https://doi.org/10.3102/978-0-935302-43-1_27

Olson, D. R. (1994). *The world on paper: The conceptual and cognitive implications of writing and reading*. Cambridge University Press.

Ong, W. J. (1982). *Orality and literacy: The technologizing of the word*. Methuen.

Plato. (1997). *Phaedrus*. In J. M. Cooper & D. S. Hutchinson (Eds.), *Plato: Complete works*. Hackett Publishing.

Spearman, C. (1904). "General intelligence," objectively determined and measured. *The American Journal of Psychology, 15*(2), 201–293.

Toulmin, S. (1990). *Cosmopolis: The hidden agenda of modernity*. University of Chicago Press.

Wegerif, R. (2007). *Dialogic education and technology: Expanding the space of learning*. Springer. This is the likely source if you intend to support the citation currently written as "Wegerif, 2013"; if so, the in-text citation should be changed to 2007.

Wegerif, R. (2025). *Rethinking educational theory: Education as expanding dialogue*. Edward Elgar Publishing.

Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. W. (2010). Evidence for a collective intelligence factor in the performance of human groups. *Science, 330*(6004), 686–688.

 © ISLS

International Society of the Learning Sciences

Designing for Augmenting Teachers' Understanding of Children with ADHD Behaviors Using a Smartwatch

Lauren Min, Chapman University, lmin@chapman.edu
Ariel Han, Utah State University, ariel.han@usu.edu

Katelyn Teav, Daisy Fernandez-Reyes, Franceli Cibrian
kteav@chapman.edu, fernandezreyes@chapman.edu, cibrian@chapman.edu
Chapman University

Abstract: This study examines expert perspectives on identifying, evaluating, and monitoring ADHD behaviors in children in the classroom. Eleven clinicians and educators participated in semi-structured interviews. Using open coding, three themes emerged: variability of ADHD behaviors, monitoring ADHD behaviors in context, and opportunities and challenges of wearables for teachers. Guided by Bronfenbrenner's Ecological Systems Theory, our findings suggest that wearable tools can act as mediators within the child's learning ecosystem, enhancing communication among teachers, clinicians, and caregivers through objective, context-aware data.

Introduction

Attention Deficit Hyperactivity Disorder (ADHD) is the most prevalent childhood psychiatric condition, affecting approximately 9.4% of children ages 2-17 in the US (Danielson et al., 2016). ADHD is characterized by differences in attention, hyperactivity, and impulsivity (APA, 2013) that challenge academic performance and peer interactions (Fried et al., 2016; Hoza, 2007). Assessing ADHD remains difficult because it depends heavily on subjective reports shaped by developmental stage, gender, cultural norms, and socioeconomic barriers (Hamed et al., 2015). Teachers are often the first to raise concerns, occupying a unique observational position to compare children's behavior against peers in a structured environment (Degroote et al., 2022). However, teachers are not qualified to diagnose ADHD, and this observational role places a heavy burden on them, requiring lengthy documentation across multiple students and leading to reporting fatigue and variability (Sherman et al., 2008). Clinicians rely on multi-setting, multi-informant data as a gold standard for diagnosis (Tamsut et al., 2024).

Researchers have begun exploring sensor-based approaches to complement traditional evaluation. Inertial sensors have demonstrated feasibility in detecting ADHD-related movement differences (O'Mahony et al., 2014)and gamified prototypes have been explored for non-invasive attentional screening (Zhang et al., 2023). Systems like MoodGems (Stefanidi et al., 2024) and CoolTaco (Silva et al., 2023) explored naturalistic behavior tracking outside of school. However, most computational approaches remain confined to controlled settings, leaving a clear gap in tools designed for real-world classroom use. This study is grounded in Bronfenbrenner's Ecological Systems Theory (EST), which frames development as shaped by nested environments, from the microsystem (family, teachers) to the macrosystem (cultural values, norms). Wearables can serve as mediating artifacts that extend teachers' observational capacity, facilitating data-informed communication across the mesosystem connecting home, school, and clinic.

To address this gap, we conducted a qualitative study with 11 clinicians and educators guided by three research questions: (1) How do experts describe the variability of ADHD behaviors across domains such as movement, speech, attention, emotion, and sensory processing? (2) How do experts determine when behaviors warrant further assessment? (3) What potential do experts see in wearable or sensor-based tools to support existing diagnostic practices?

Methods

Eleven participants (6 clinicians, 5 educators) were recruited via email and flyers based on experience working with children with ADHD or other neurodevelopmental disabilities. Clinicians included active practitioners and those with prior clinical experience; educators had 6-14 years of classroom experience supporting students with ADHD. Interviews of 40 semi-structured questions were conducted in person or via Zoom, covering ADHD symptom identification, behavioral monitoring, movement and speech differences, sensor use concerns, and useful data visualizations. Each interview was recorded, transcribed, and analyzed using open coding (Khandkar, 2009), with codes organized collaboratively on a shared Miro board to surface thematic trends (Castleberry & Nolen, 2018).

 © ISLS

Findings

Variability of ADHD behaviors in the classroom

Experts consistently identified movement, speech, attention, emotion, and sensory processing as the primary behavioral domains distinguishing students with ADHD from neurotypical peers. Across all five domains, behavior was characterized not by quantity but by quality: movement tended to be sporadic and disorganized rather than merely excessive; speech involved inappropriate loudness and difficulty with turn-taking rather than simple talkativeness; and attention oscillated between distractibility and hyperfocus, with educators most concerned about the former during sustained tasks. Emotional dysregulation manifested as heightened reactivity that created barriers to peer interaction, while sensory differences often presented as movement-seeking behaviors that educators misread as ordinary fidgeting. This multidimensional variability underscores ADHD as a spectrum rather than a uniform profile, and reflects how child-level characteristics interact dynamically with the immediate classroom microsystem, including teacher expectations, peer dynamics, and task demands.

Monitoring ADHD Behaviors in the Classroom

Experts noted that behaviors become clinically significant when they exceed developmental expectations, impair learning or participation, and persist across time and contexts. A central finding was the contextual dependence of ADHD behaviors: what is disruptive in a structured classroom may be unremarkable at home, and all children exhibit disruptive behaviors at some point. This underscores the need for longitudinal, multi-informant assessments that integrate contextual information across settings. Experts also cautioned that interpretations are susceptible to bias rooted in cultural norms and gender expectations, as ADHD assessment has historically centered white, middle-class male presentations, leaving many children underidentified. Girls in particular are socialized to mask symptoms, and teachers must remain alert to hidden compensatory behaviors in students who appear compliant or quiet yet are experiencing significant internal distress. From an EST perspective, these dynamics span the mesosystem and macrosystem, shaping how teachers perceive and respond to children's behavior in ways that can perpetuate inequities in identification.

Opportunities and Challenges for Wearables in Supporting Teachers

Experts saw significant potential in smartwatches to provide objective, continuous, and context-sensitive data on ADHD behaviors, capturing variability across tasks, times of day, and weeks in ways that episodic observations cannot. Data of interest included frequency and quality of movement, voice modulation levels, and behavioral patterns across different classroom contexts such as structured lessons versus recess. Smartwatches were seen as less disruptive than tablets or cameras, and their data could be exported at varying levels of granularity to suit different stakeholder needs. Engaging children in self-tracking was also viewed as promising for motivation and compliance. However, experts cautioned that wearables must align with practical classroom realities: concerns included device removal, discomfort, altered behavior due to monitoring awareness, and the risk of data overload. Wearable summaries must augment rather than replace teachers' observation logs, be culturally contextualized, and integrate concisely into existing workflows. When thoughtfully designed, such tools function as mediating artifacts within the mesosystem, fostering shared understanding among teachers, clinicians, and caregivers.

Discussion and Conclusion

This study explored how clinicians and educators interpret ADHD behaviors in classrooms and how smartwatch-based tools could augment this process. Our findings confirm that ADHD identification is an inherently ecological process shaped by child characteristics, teacher perceptions, family context, and broader sociocultural norms. Technologies designed to support educators must model individual baselines, adapt to contextual shifts, and enhance cross-system communication rather than imposing uniform detection thresholds. Experts emphasized that wearable data is most valuable when it surfaces behavioral variability across contexts, supports collaboration among stakeholders, and is presented in formats that complement rather than complicate teachers' existing practices. These findings suggest three design directions: (1) Model contextual variability across tasks, routines, and social settings; (2) Support cross-system collaboration through glanceable, shareable summaries aligned with teachers' workflows; and (3) Promote equity and cultural sensitivity by actively challenging biases embedded in existing diagnostic practices. This work extends EST into educational technology by framing wearables as context-aware learning supports that connect data-driven insights with teachers' pedagogical judgment, ultimately strengthening the ecosystem that surrounds and supports children with ADHD.

 © ISLS

References

American Psychiatric Association, DSM-5 Task Force. (2013). Diagnostic and statistical manual of mental disorders: DSM-5™ (5th ed.). American Psychiatric Publishing, Inc.. https://doi.org/10.1176/appi.books.9780890425596

Danielson, M. L., Bitsko, R. H., Ghandour, R. M., Holbrook, J. R., Kogan, M. D., & Blumberg, S. J. (2018). Prevalence of Parent-Reported ADHD Diagnosis and Associated Treatment Among U.S. Children and Adolescents, 2016. *Journal of Clinical Child & Adolescent Psychology*, 47(2), 199–212. https://doi.org/10.1080/15374416.2017.1417860

Darling, N. (2007). Ecological Systems Theory: The Person in the Center of the Circles. Research in Human Development, 4(3–4), 203–217. https://doi-org.chapman.idm.oclc.org/10.1080/15427600701663023

Degroote, E., Brault, M. C., & Van Houtte, M. (2022). Suspicion of ADHD by teachers in relation to their perception of students' cognitive capacities: do cognitively strong students escape verdict? *International Journal of Inclusive Education*, 28(9), 1740–1754. https://doi.org/10.1080/13603116.2022.2029962

Fried, R., Petty, C., Faraone, S. V., Hyder, L. L., Day, H., & Biederman, J. (2016). Is ADHD a Risk Factor for High School Dropout? A Controlled Study. *Journal of attention disorders*, 20(5), 383–389. https://doi.org/10.1177/1087054712473180

Hamed, A. M., Kauer, A. J., & Stevens, H. E. (2015). Why the Diagnosis of Attention Deficit Hyperactivity Disorder Matters. *Frontiers in Psychiatry*, 6(168). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4659921

O'Mahony, N., Florentino-Liano, B., Carballo, J. J., Baca-García, E., & Rodríguez, A. A. (2014). Objective diagnosis of ADHD using IMUs. Medical Engineering & Physics, 36(7), 922–926. https://doi.org/10.1016/j.medengphy.2014.02.023

Silva, L. M., Cibrian F. L., Monteiro, E. M., Bhattacharya, A., Beltran, J. A., Clarisse Bonang, Epstein, D. A., Schuck, S., Lakes, K. D., & Hayes, G. R. (2023). Unpacking the Lived Experiences of Smartwatch Mediated Self and Co-Regulation with ADHD Children. https://doi.org/10.1145/3544548.3581316

Sherman, J., Rasmussen, C., & Baydala, L. (2008). The impact of teacher factors on achievement and behavioural outcomes of children with Attention Deficit/Hyperactivity Disorder (ADHD): a review of the literature. *Educational Research*, 50(4), 347–360. https://doi.org/10.1080/00131880802499803

Stefanidi, E., Wassmann, J. L. B., Woźniak, P. W., Spellmeyer, G., Rogers, Y., & Niess, J. (2024). MoodGems: Designing for the Well-being of Children with ADHD and their Families at Home. *Proceedings of the 23rd Annual ACM Interaction Design and Children Conference*, 4, 480–494. https://doi.org/10.1145/3628516.3655795

Tamsut, A., Asadi, H., Sinai, G. N., Saka, N., & Pollak, Y. (2024). Elementary School Teachers' and Counselors' Decisions on Referring Students for Evaluation: The Impact of ADHD Traits, Achievement, and Giftedness. Child Psychiatry & Human Development. https://doi.org/10.1007/s10578-024-01777-0

Zhang, H., Liu, Y., Zhu, S., & Fang, X. (2023). Prototype Design of Wearable Multimodal Interaction Based on Preliminary Screening for ADHD in Children in Preschool Age. https://doi.org/10.1145/3585088.3593898

Acknowledgment

This work was partially supported by the National Science Foundation (NSF) under Grant #2245495 awarded to Franceli L. Cibrian. Any opinions, findings, and conclusions are those of the authors and do not necessarily reflect the views of NSF. We also thank our participants, collaborators, and students for their contributions to the broader research efforts. This research was approved by University IRB-24-31.

International Society of the Learning Sciences

"He's Gonna Say It's Trash": Students' Emotional Experiences with AI Feedback in Collaborative Scientific Inquiry

Megan Humburg, Indiana University, mahumbur@iu.edu
Namrata Srivastava, Vanderbilt University, namrata.srivastava@vanderbilt.edu
Sarah Burriss, Vanderbilt University, sarah.burriss@vanderbilt.edu
Yeo Jin Kim, North Carolina State University, ykim32@ncsu.edu
Joshua Danish, University of Illinois Urbana-Champaign, jdanish@illinois.edu
Cindy E. Hmelo-Silver, Indiana University, chmelosi@iu.edu
Shruti Jain, Vanderbilt University, shruti.jain@vanderbilt.edu
Clayton Cohn, Vanderbilt University, clayton.a.cohn@vanderbilt.edu
Umesh Timalsina, Vanderbilt University, umesh.timalsina@vanderbilt.edu
Krista Glazewski, North Carolina State University, kdglazew@ncsu.edu
Gautam Biswas, Vanderbilt University, gautam.biswas@vanderbilt.edu
James Lester, North Carolina State University, lester@ncsu.edu

Abstract: This study explores how 5th grade students playing a scientific inquiry game interact with (and react to) AI-generated feedback presented by in-game agent characters. We explore how AI conversational agents impact student inquiry and emotional engagement as students build socioscientific arguments. Results indicate that while students initially take up agent feedback to improve their argumentation, prolonged cycles of revision lead to frustration and negative reflections as students struggle to utilize AI feedback.

Introduction and theoretical framework

Across classrooms, many educators are exploring how to integrate the latest advances in generative AI. Early evidence highlights that AI feedback could enhance student engagement and motivation in learning games (Mehrvarz et al., 2025) and that AI conversational agents can support students in monitoring and reflecting on their learning (Järvelä et al., 2025). However, a critical aspect of this integration lies in understanding how interaction with AI agents affects students' social and emotional wellbeing. Emotions play a central role in shaping how and what students learn (Jaber & Hammer, 2016), so it is vital to understand how AI agents impact learners' affective experiences. This study explores the influence of AI feedback for student argumentation on 5th grade learners' emotional experiences during game-based scientific inquiry. Receiving feedback is an emotion-laden process in which students may perceive critical feedback as a threat to their skills and identities. Trust and relationship-building are viewed as core components of how students respond to and internalize feedback (Molloy et al., 2019). This highlights the importance of understanding how students respond to feedback that is offered from a non-human source, such as generative AI, that lacks such social abilities. Our poster will explore how students react to AI feedback on their writing during a collaborative, playful learning activity, and how these emotional reactions appear to influence students' engagement with scientific inquiry.

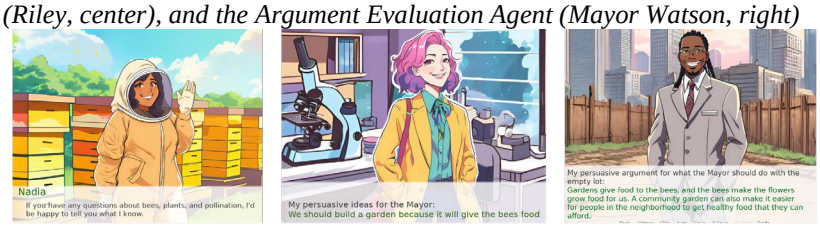
Design

This study analyzes a pilot implementation of SciStory: Pollinators, a 2D visual novel-style game designed to engage middle-school students in open-ended exploration of a socioscientific dilemma: *What should a community do with an empty plot of land?* As students visit different locations, converse with multiple characters, and take notes in their in-game notebooks, they work in pairs to construct a persuasive argument for the mayor, advocating to convert the empty lot into either a community garden or a parking structure. Through these interactions, students learn about the relationship between pollinators and people and the role of community food systems. The game integrates AI agents embedded within multiple narrative characters to support scientific inquiry (see Figure 1). Content Knowledge Agents provide information about pollinators and human-environment interactions. Argument Feedback Agents offer feedback on students' draft arguments, prompting students to strengthen claims and incorporate evidence. Finally, the Argument Evaluation Agent reviews students' final arguments and determines whether they are persuasive enough to influence the community. All agents were powered by a locally hosted Llama-3.1-8B model using a retrieval-augmented generation (RAG) pipeline grounded in a curated science knowledge base and response templates co-designed with educators. Student input is sorted (question, evaluation, or inappropriate language), and then agents use tailored prompts with gameplay information and argument history dynamically filled in to generate responses.

 © ISLS

Figure 1

A Content Knowledge Agent (Nadia the beekeeper, left), an Argument Feedback Agent (Riley, center), and the Argument Evaluation Agent (Mayor Watson, right)



Methods

Three 5th-grade classrooms in a rural Midwestern U.S. school participated, with 15 dyads (30 students) in which both students consented to have their data used. Each classroom engaged with SciStory over two 60-minute sessions for two days. Laptop screen recordings, audio recordings, and log files with gameplay data were collected to analyze how students interacted with the agents. On Day 1, students familiarized themselves with the story and core dilemma, and on Day 2 they began interacting with the AI agents to share their arguments and receive feedback. To analyze students' reactions to agent feedback, data was segmented into "agent episodes" that began with a student query to the agent and ended when the next query was submitted. This allowed us to track ensuing conversations and how the response impacted students' subsequent inquiry. We conducted a qualitative content analysis of audio and screen data, triangulated with log data, to identify and refine themes regarding (1) the emotions students expressed in response to agent feedback, and (2) the impact of these emotions on students' engagement or disengagement with the learning activity.

Results

Frustration was a key emotion that students commonly expressed when receiving AI feedback, particularly 1) when they struggled to use game tools to compose an argument to share with the agent, and 2) when the agent was perceived as harsh or repetitive in their judgments and suggestions. The first source of frustration appeared to stem from difficulties with spelling and typing as well as some technical issues, which slowed down student progress. The second source of frustration compounded on top of the first, because after drafting arguments to share with the Mayor agent, students would often receive feedback that their argument was either too vague or simplistic. This feedback was often accurate but sometimes included hallucinated suggestions for irrelevant evidence, which amplified student confusion. Initially, many groups responded to feedback by speaking with more characters and gathering new notes to improve arguments. However, when groups revised their arguments and again received feedback that they were not yet persuasive enough, some dyads reacted to the feedback with harsh assessments (e.g., "Elliot, you're useless" in response to repetitive advice given by an argument feedback agent) or exasperated exclamations such as, "I don't know how to get freaking evidence. I'm gonna crash out on you, do you understand that?" Some students attempted to resolve their frustration by communicating it, such as when one group said to Riley, "I found out [the Mayor] is so mean." However, the agent was unable to respond helpfully to this input and instead continued to offer argumentation advice. Students also began to negatively reflect on feedback before it was even received, with one group guessing, "He's gonna say it's trash" while waiting for the Mayor's response. Overall, these initial results, which we will explore further in our poster, highlight ways in which AI agent feedback can lead to frustration and disengagement when agents are unable to meaningfully support the emotional layers of receiving and responding to feedback.

Discussion and conclusion

The significant presence of student frustration in our data suggests that the design of AI conversational agents must take into account the emotional layers of receiving feedback in the delivery of critiques on student writing. This supports recent studies that have highlighted the importance of pedagogical agents providing emotional feedback to support student motivation and reduce confusion (Lang et al., 2022). The goal is not to eliminate frustration, but rather to design agents that productively respond to and scaffold the regulation of challenging emotions, while providing feedback that is perceived as supportive rather than harsh or overwhelming. However, much is left to learn about how to weave together emotional and cognitive feedback when presenting students with AI-generated guidance, as well as learners' overall perceptions of receiving such feedback from AI tools versus human educators (Le et al., 2025). Our presentation will explore what aspects of AI feedback encouraged students to deepen their inquiry and where teacher feedback is valuable to balance the frustration and confusion students experience when engaging in multiple cycles of sharing and feedback with AI agents.

References

Jaber, L. Z., & Hammer, D. (2016). Engaging in Science: A Feeling for the Discipline. *Journal of the Learning Sciences, 25*(2), 156-202. https://doi.org/10.1080/10508406.2015.1088441

Järvelä, S., Rummel, N., Cukurova, M., Radtke, A., Meyer, J., Yang, S., ... & Papadopoulos, P. M. (2025). Exploring Human-AI Interactions Across Various Learning Tasks: Diverse Perspectives on Augmenting Learner. In *Proceedings of the 18th International Conference on Computer-Supported Collaborative Learning-CSCL 2025*, pp. 481-489. International Society of the Learning Sciences.

Lang, Y., Xie, K., Gong, S., Wang, Y., & Cao, Y. (2022). The Impact of Emotional Feedback and Elaborated Feedback of a Pedagogical Agent on Multimedia Learning. *Frontiers in psychology*, *13*, 810194. https://doi.org/10.3389/fpsyg.2022.810194

Le, H., Shen, Y., Li, Z., Xia, M., Tang, L., Li, X., ... & Fan, Y. (2025). Breaking human dominance: Investigating learners' preferences for learning feedback from generative AI and human tutors. *British Journal of Educational Technology*, 56, 1758–1783. https://doi.org/10.1111/bjet.13614

Mehrvarz, M., McLaren, B. M., Qiu, A., Purtell, K., Yesilyurt, S., & Nguyen, D. V. (2025). ChatGPT as a Feedback Provider in Digital Learning Games: A Systematic Literature Review. In *Proceedings of the 19th International Conference of the Learning Sciences-ICLS 2025*, pp. 1749-1753. International Society of the Learning Sciences.

Molloy, E., Noble, C., & Ajjawi, R. (2019). Attending to Emotion in Feedback. In *The Impact of Feedback in Higher Education: Improving Assessment Outcomes for Learners* (pp. 83-105). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-25112-3_6

Acknowledgments

This material is based upon work supported by the National Science Foundation under Award No. DRL-2112635. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

International Society of the Learning Sciences

Iterative Design for Distributed Embodied Collaboration

Casey Smith, Mike Tissenbaum
cjsmith0@illinois.edu, miketiss@illinois.edu
University of Illinois Urbana-Champaign

Abstract: To effectively support remote maker sessions, tools are needed that allow distributed embodied collaborations to emerge naturally rather than forcing conformity to existing video conferencing. We present an overview of a multi-year iterative design, testing, and evaluation process of REACH (Remote Embodiment for Augmented Collaborative Help), offering insights into developing similar systems to support distributed embodied collaborations.

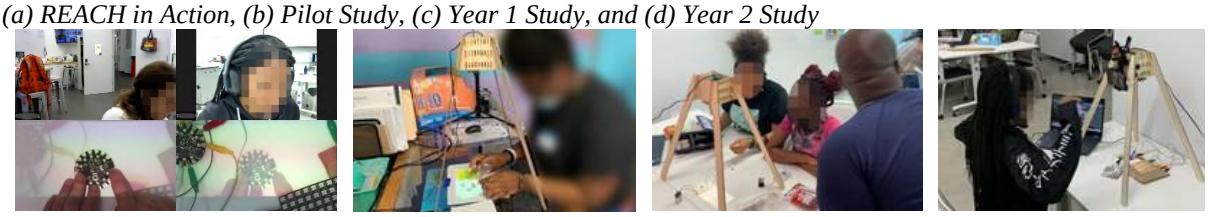
Introduction

Collaboration between learners at a distance has long been an interest in the learning sciences (Dillenbourg et al., 1996); however, the ways typical technologies mediate these exchanges can make activities like remote making difficult (Jayathirtha et al., 2020). During co-located exchanges participants use embodied gestures around physical artifacts to guide attention (Goodwin, 2003; McNeill, 1992), communicate shared meaning without domain-specific language (Roth, 2001), and help cue learning (Lindgren, 2014). However, video calls often limit embodied interaction with physical objects, making explanation, debugging, and collaboration more difficult (Licoppe et al., 2017).

In response, REACH is a low-cost camera–projector system that merges remote workspaces by overlaying projected gestures and artifacts to enable real-time interaction around shared objects (Smith & Tissenbaum, 2022). Each participant's workspace is captured and projected onto their partner's surface, allowing users to gesture around and manipulate artifacts as if co-located (Figure 1a). Developed through iterative design-based research in makerspace camps, REACH has evolved across multiple cycles to improve functionality and explore distributed embodied collaboration during authentic making practice.

The first REACH prototype, developed in early 2020, was tested in a remote circuit-building activity between a mentor and student. Early observations suggested its potential for remote gesturing but also revealed three key design limitations: time-consuming manual camera–projector alignment, visual artifacts from overlapping camera-protector views, and limited networking support. To address these issues, we developed calibration software to improve alignment between physical and projected objects and migrated from *WebRTC* (n.d.) to *Twilio* (n.d.) for more reliable connectivity and recording. A web-based dashboard was also introduced to support remote setup and debugging. Attempts to mitigate visual echo artifacts in software (Liu et al., 2024), proved computationally intensive and ultimately unnecessary under typical lighting conditions. With these design updates, we piloted REACH in an afterschool program using a short puzzle activity (Figure 1b), where pairs of students used gestures to coordinate pieces across locations. While students successfully collaborated to solve the puzzle across sites they noted that face-to-face communication would further improve their interactions.

Figure 1
(a) REACH in Action, (b) Pilot Study, (c) Year 1 Study, and (d) Year 2 Study



(a) (b) (c) (d)

Year one – building and debugging circuits with remote peers

Adding a *Zoom* (n.d.) session at each station, we deployed REACH in a jigsaw activity during a 2022 study, where two pairs of students collaborated remotely to assemble and debug a circuit (Figure 1c). Using the Divergent Collaborative Learning Mechanisms (DCLM) framework (Tissenbaum et al., 2017) to analyze 12 cases, we showed that REACH extends key affordances of co-located making into distributed settings, enabling effective remote peer collaboration (Smith & Tissenbaum, 2025). Participants reported that, "it was easy to transfer skill" but noted that "most communication happened through talk" due to the quality of the projection. The projectors' low resolution and small image size made circuit details difficult to see and limited participation from local

 © ISLS

partners. Additionally, using a Chromebook for Zoom required headphones in the noisy camp environment, effectively excluding the second student on each side.

Year two – supporting making through remote mentorship

In the second iteration in 2023, students engaged in open-ended maker projects while receiving just-in-time support from remote mentors to troubleshoot, refine, and extend their designs. Based on feedback from year one, we improved the REACH Dashboard so sessions could be launched directly from the web interface, removing the need to start each projector manually, and simplifying setup in the field. Zoom sessions were shifted to iPhones with Bluetooth headsets making it easier for students to present their projects to remote mentors (Figure 1d), but remained cumbersome due to complex recording settings, account management, and frequent accidental disconnects. Using DCLM to examine 11 cases, we showed that REACH supports remote mentorship by enabling embodied multimodal interaction, giving rise to distinct mentoring patterns such as guidance, brainstorming, and debugging (Smith & Tissenbaum, *in review a*). Users again noted resolution issues, though improved lighting and larger project scales made these less disruptive. Students also struggled to fit larger projects under the projector and share laptop screens when seeking programming help. While they worked around these limitations using Zoom, the challenges highlighted the need for a more flexible stand design and integrated screen sharing.

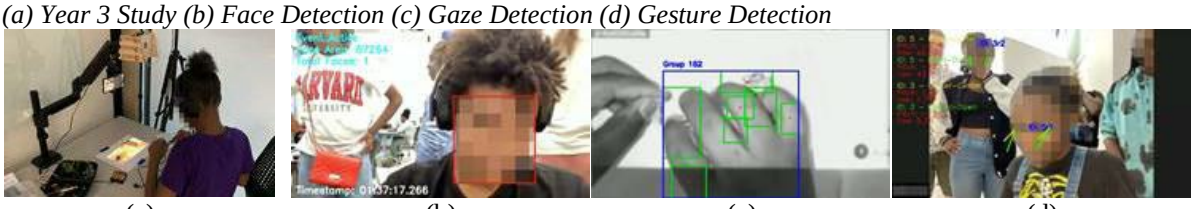
Year three – shifting roles in distributed embodied collaboration

In the 2024 offering, students participated in a distributed maker camp, collaborating with remote peers and mentors to build and share projects while shifting between roles across different interaction types. Lessons from Year 2 informed us of two key improvements: custom video conferencing and a redesigned stand. We developed REACH Observer, a two-way video app integrated with the projector's Twilio backend to streamline setup, recording, and post-processing. Audio clarity remained challenging in the noisy camp environment, leading to a switch to noise-cancelling headsets. This improved clarity, but if worn backwards would mute students. Labeling helped, but additional training and automated prompts for low microphone volume could improve reliability. To better support larger projects, we redesigned the stand using a VESA-compatible mount, improving stability, image size, and workspace access (Figure 2a). This enabled students to share larger artifacts, including full laptop screens, within the projected workspace.

Ten REACH stations across two camps and a remote mentor site generated over 100 GB of data per day with multiple video channels per link supporting the "always-on" interaction. To manage this dataset (3.7 TB total), we used computer-vision tools (Bradski, 2000) to help automate segmentation. Face detection identified active sessions (Figure 2b) and motion tracking isolated worksurface gestures (Figure 2c). However, gaze estimation (Figure 2d) using MediaPipe (Lugaresi et al., 2019), proved unreliable as students' eyes were often out of frame, requiring manual coding. Analysis of 10 cases using DCLM, showed that participant roles shaped the use of DCLM, with mentor–student interactions emphasizing boundary-spanning action and coordinated joint attention, and peer interactions relying more on boundary spanning perception to coordinate peer visibility and recognition (Smith & Tissenbaum, *in review b*).

Figure 2

(a) Year 3 Study (b) Face Detection (c) Gaze Detection (d) Gesture Detection



(a) (b) (c) (d)

Conclusion

This paper presents the design-based research journey of REACH over several years, highlighting the importance of iterative design for effectively supporting learners distributed embodied collaboration with physical artifacts, even when they are not co-located. These kinds of design stories are critical for the fields of computer-supported collaborative learning and the learning sciences broadly, as they outline the often messy and non-linear progression of technology and theory (Kali, 2016). By extending the benefits of co-located makerspaces into distributed contexts, the design trajectory of REACH presented here is intended to offer a possible touchpoint for researchers investigating new tools for distributed embodied collaborations that offer a new avenue of support for hybrid and remote education grounded in constructionist, sociocultural, and embodied learning theories.

References

Bradski, G. (2000). The opencv library. *Dr. Dobb's Journal: Software Tools for the Professional Programmer*, *25*(11), 120–123.

Dillenbourg, P., Baker, M., Blaye, A., & O'malley, C. (1996). The evolution of research on collaborative learning In H. Spada and P. Reimann (Eds) Learning in Humans and Machines. *Elsevier*, *1*(1), 58–94.

Goodwin, C. (2003). The Body in Action. In J. Coupland & R. Gwyn (Eds.), *Discourse, the Body, and Identity* (pp. 19–42). Palgrave Macmillan UK.

Jayathirtha, G., Fields, D., Kafai, Y. B., & Chipps, J. (2020). Supporting making online: The role of artifact, teacher and peer interactions in crafting electronic textiles. *Information and Learning Sciences, 121*(5/6), 381–390.

Kali, Y. (2016). *Transformative Learning in Design Research: The Story Behind the Scenes*. 20.

Licoppe, C., Luff, P. K., Heath, C., Kuzuoka, H., Yamashita, N., & Tuncer, S. (2017). Showing Objects: Holding and Manipulating Artefacts in Video-mediated Collaborative Settings. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 5295–5306). Association for Computing Machinery. 8

Lindgren, R. (2014). Getting into the cue: Embracing technology-facilitated body movements as a starting point for learning. In *Learning Technologies and the Body* (pp. 51–66). Routledge.

Liu, J., Smith, C. J., & Tissenbaum, M. (2024). REACH Projector: A Remote Collaboration Projector-Camera System with an Effective Visual Echo Cancellation Algorithm. *Companion Publication of the 2024 Conference on Computer-Supported Cooperative Work and Social Computing, CSCW Companion '24*, 175–178.

Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., Zhang, F., Chang, C.-L., Yong, M. G., Lee, J., Chang, W.-T., Hua, W., Georg, M., & Grundmann, M. (2019). *MediaPipe: A Framework for Building Perception Pipelines* (arXiv:1906.08172). arXiv.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago press.

Roth, W.-M. (2001). Gestures: Their role in teaching and learning. *Review of Educational Research, 71*(3), 365–392.

Smith, C., & Tissenbaum, M. (2022). REACH Projector: Remote Embodiment for Augmented Collaborative Help. *Proceedings of the 15th International Conference on Computer-Supported Collaborative Learning*, 336–339.

Smith, C., & Tissenbaum, M. (2025). REACH: Extending reality for distributed collaborative making. *Computers & Education: X Reality*, *7*, 100111.

Smith, C., & Tissenbaum, M. (*in review a*). REACH : Supporting Multimodal Interactions for Distributed Making Mentorship. *IJCCI*.

Smith, C., & Tissenbaum, M. (*in review b*). REACH: Comparing Remote Interactions in a Distributed Makerspace. *IJCSCL*.

Tissenbaum, M., Berland, M., & Lyons, L. (2017). DCLM framework: Understanding collaboration in open-ended tabletop learning environments. *International Journal of Computer-Supported Collaborative Learning*, *12*(1), 35–64.

Twilio. (n.d.). [Computer software]. Retrieved https://www.twilio.com/

WebRTC. (n.d.). [Computer software]. Retrieved https://webrtc.org/

Zoom. (n.d.). [Computer software]. Retrieved https://www.zoom.com/

Acknowledgments

This research is supported by grants from the NSF (2048833) and the Institute for Inclusion, Diversity, Equity & Access at the University of Illinois Urbana-Champaign (GIANT2021-04)

International Society of the Learning Sciences

Complementary Roles of Collaborative Reflection and AI-generated Feedback to Support Students' Science Writing

Indrani Dey, University of Wisconsin–Madison, idey2@wisc.edu
ChanMin Kim, Pennsylvania State University, cmk604@psu.edu
Sadhana Puntambekar, University of Wisconsin–Madison, puntambekar@education.wisc.edu

Abstract: Automated assessment can provide timely feedback on students' writing, but students often struggle to interpret and apply feedback effectively. This study examined the extent to which collaborative reflection activities can support students to interpret, evaluate, and act on automated feedback to improve their scientific explanations.

Background and motivation

AI-based assessment of student writing can provide timely feedback (Celik et al., 2022), but research shows that relying on such feedback alone may limit students' metacognitive engagement (Besigomwe, 2025), self-regulation (Fan et al., 2025), and ability to interpret and apply feedback effectively (Dey et al., 2024; Puntambekar et al., 2023). Thus, AI feedback by itself is often insufficient to support deep conceptual understanding or meaningful revision without complementary human-centered support(s). Collaborative reflection and discourse offer such support by engaging students in cognitive (Kollar & Fisher, 2010) and regulatory (Panadero et al., 2015) processes that support their metacognitive development (Wiedbusch et al., 2022). By critically evaluating peers' work, students can identify gaps in understanding, recognize strengths, and consider alternative approaches (Gerard & Linn, 2022). Integrating AI-generated feedback with collaborative reflection can create a synergistic human-AI partnership, in which AI provides diagnostic input and collaborative reflection scaffolds interpretation, evaluation, and application of feedback, thus forming an ecosystem of support (Puntambekar et al., 2024).

While prior studies have examined AI-based feedback and collaborative reflection, less is known about the extent to which collaborative reflection serves as a metacognitive scaffold that supports students' use of AI-generated feedback for revision. Our study explores the extent to which collaborative reflection activities support students in meaningfully revising their scientific explanations. Our research questions were:

1. *What issues do students identify and discuss in sample essays during collaborative reflection?*
2. *To what extent do students' reflection of AI-generated feedback translate into revisions in their own essays?*

Methods

Study context and participants

Students (*n=31*) from two classes taught by the same teacher at a Midwestern US school participated in this study. Students engaged in a design-based physics unit on motion, forces, and energy, conducting virtual experiments and writing essays in a digital notebook. Once students submitted their essays, PyrEval (Gao et al., 2019; Passonneau et al., 2018), an NLP tool, evaluated them for the presence of six science concepts. PyrEval uses a wise-crowd model, comparing student writing to expert-generated reference essays by constructing weighted vector representations of concepts and their similarity. The automated feedback indicated whether each concept was detected or not in each essay. To support automated feedback interpretation, we implemented participatory structures (Tabak & Baumgartner, 2004), including collaborative group reflection (Kollar & Fischer, 2010) and teacher guidance. Students collaboratively reviewed three anonymous sample essays with PyrEval feedback and noted, in individual reflection worksheets, whether ideas were correctly identified, why they were detected or missed, and how the writing could be improved. Students then revised their own essays using automated feedback.

Data Sources and Analysis

Data came from three sources: (ii) students' reflection worksheets, (ii) approximately 125 minutes of video across six groups during their collaborative reflection activity, and (iii) students' initial and revised essays. We inductively and deductively identified four themes from students' discussion and annotations of sample essays in their reflection worksheets: *(i) missing ideas, (ii) evidence to support claims, (iii) refinement of explanations, and (iv) writing quality*. Each theme was coded as 1 (identified) or 0 (not identified).

We then applied the same four themes to examine whether students addressed these same issues in their own essays during revision. Revisions were coded as 1 (*addressed*, i.e, one or more ideas needed to be addressed

 © ISLS

and at least one idea was addressed), –1 (*did not address*, i.e., one or more ideas needed to be addressed but none were addressed), or 0 (*did not need to be addressed*).

Results and Discussion

Issues *identified* by students during collaborative reflection

We identified four recurring issues from student reflection worksheets and group discussions. The reflection worksheets showed that all students identified *missing ideas* (100%), and most noted the *need to add evidence* (87.1%), *improve explanations* (96.8%), and *overall writing quality* (87.1%) (see Figure 1a). The excerpts below illustrate how students discussed these issues during collaborative reflection.

Table 1

Excerpts Illustrating How Students Identified a) Ideas and b) Potential Issues in Sample Essays

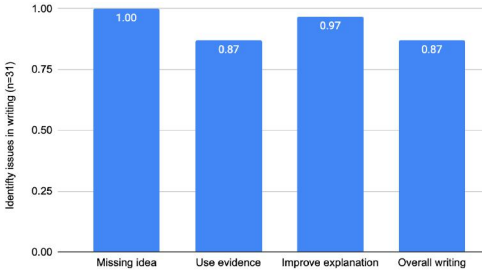
(a) identifying ideas in the sample essay	
S1: So right off of the bat, it talks about height and potential energy… Does anyone else think that..	
S2: This first sentence explains the height and the PE. [*reads*] "*When you have a higher hill height, you get more PE*"... and the second sentence says it too.	
(b) identifying the potential issues in the sample essay	
(i) missing ideas	***(iii) improve explanations***
S1: Okay, total energy… does it even mention total energy? S2: Nope. S3: Is there even any total energy? S4: No. S1: No, they don't mention it.	S1: Okay, mass and energy.. Uhh… they really just didn't, like, talk about how it affects, like, the energy.. really… S2: It just said like what they were doing in their roller coaster and not like.. S1: Mmhmm. S3: Yeah.
(ii) lack of evidence to support claims	***(iv) improve writing quality***
S1: Would you say they had like.. Reasoning? Like evidence to back it up? S2: Ummm.. like they never used evidence at all. S3: Mmhmm.. Yup.. I mean, they just said like what they wanted to do… But never actually explained some evidence.	S1: They kind of like talk in circles...like this part where they said [*pointing at their worksheet and reading*] "*you get more potential energy because there is more potential energy for energy.. because the hill is higher*". S2: They're kind of restating, like, the same thing, for the first one.

Issues *addressed* by students in their own essays

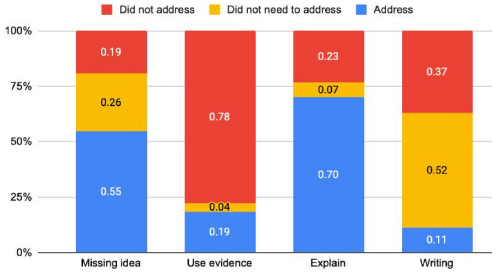
We examined the proportion of students who identified issues in sample essays (Figure 1a) to those who addressed them in their revisions (Figure 1b). Students were more successful at identifying issues in others' writing than addressing them during their own revisions. The most common revisions involved refining explanations (*0.70*) and adding missing ideas *(0.55*), suggesting students were more comfortable elaborating on existing ideas or incorporating identified missing ideas. Although most students recognized the need for supporting evidence in sample essays (*0.87*), many did not add evidence to their own writing when needed (*0.78*). Similarly, while students readily identified writing quality issues in others' essays, applying these insights to their own writing remained challenging, with about one-third (*0.37*) failing to revise clarity or repetition issues.

Figure 1

(a) Proportion of Students Who Identified Issues in the Sample Essays



(b) Proportion of Students Who Identified Issues but Addressed, Did Not Address, or Did Not Need to Address Them During Revision



Overall, our results indicate that collaborative reflection helped students identify issues in the example essays they reviewed together. However, this awareness did not always successfully translate into their own revisions, suggesting a need for additional support to help students apply reflection into effective revisions.

 © ISLS

References

Besigomwe, K. (2025). Algorithmic self-deception: How AI-generated feedback skews learners' self-reflection. *Cognizance Journal of Multidisciplinary Studies*, *5*(7), 333-342.

Celik, I., Dindar, M., Muukkonen, H., & Järvelä, S. (2022). The promises and challenges of artificial intelligence for teachers: A systematic review of research. *TechTrends*, *66*(4), 616-630.

Dey, I., Gnesdilow, D., Passonneau, R., & Puntambekar, S. (2024). Potential pitfalls of false positives. In *International Conference on Artificial Intelligence in Education* (pp. 469-476). Cham: Springer Nature Switzerland.

Fan, F., Wang, S., Dinuer, M., Hemuti, M., Nie, X., & Yang, L. T. (2025). Enhancing Students' Metacognition with Innovative IA-Based Metacognitive Reflective Learning Tool. In *IEEE Transactions on Learning Technologies*, *18*, 699-715. doi: 10.1109/TLT.2025.3580536

Gao, Y., Sun, C., & Passonneau, R. J. (2019). Automated pyramid summarization evaluation. *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)*, 404–418.

Gerard, L., Linn, M. C., & Berkeley, U. C. (2022). Computer-based guidance to support students' revision of their science explanations. *Computers & Education*, *176*, 104351.

Kollar, I., & Fischer, F. (2010). Peer assessment as collaborative learning: A cognitive perspective. *Learning and Instruction*, 20(4), 344-348.

Panadero, E., Kirschner, P. A., Järvelä, S., Malmberg, J., & Järvenoja, H. (2015). How individual self-regulation affects group regulation and performance: A shared regulation intervention. *Small Group Research*, *46*(4), 431-454.

Puntambekar, S., Dey, I., Gnesdilow, D., Passonneau, R. J., & Kim, C. (2023). Examining the effect of automated assessments and feedback on students' written science explanations. In *Blikstein, P., Van Aalst, J., Kizito, R., & Brennan, K. (Eds.), Proceedings of the 17th International Conference of the Learning Sciences - ICLS 2023* (pp. 1865-1866). International Society of the Learning Sciences.

Puntambekar, S., Gnesdilow, D., Passonneau, R., & Kim, C. (2024). AI-Human Partnership to Help Students Write Science Explanations. In *Lindgren, R., Asino, T. I., Kyza, E. A., Looi, C. K., Keifert, D. T., & Suárez, E. (Eds.), Proceedings of the 18th International Conference of the Learning Sciences - ICLS 2024* (pp. 234-241). International Society of the Learning Sciences.

Passonneau, R. J., Poddar, A., Gite, G., Krivokapic, A., Yang, Q., & Perin, D. (2018). Wise crowd content assessment and educational rubrics. *International Journal of Artificial Intelligence in Education*, *28*(1), 29–55.

Tabak, I., & Baumgartner, E. (2004). The teacher as partner: Exploring participant structures, symmetry, and identity work in scaffolding. *Cognition and Instruction*, *22*, 393–429.

Wiedbusch, M. D., Sonnenfeld, N., Dever, D., & Azevedo, R. (2022). Clustering learner's metacognitive judgment accuracy and bias to explore learning with AIEd systems. In M. M. Rodrigo, N. Matsuda, A. I. Cristea, & V. Dimitrova (Eds.), *Artificial Intelligence in Education* (Vol. 13355, pp. 402–413). Springer International Publishing. https://doi.org/10.1007/978-3-031-11644-5_33

Acknowledgments

We thank the teacher and students who participated in this study. This research has been supported by a DRL grant from the National Science Foundation (Grant# 2010483).

International Society of the Learning Sciences

Collaborative Virtual Reality Worlds, Cultural Play: Re-imagining Game Design through Interactional Tool Exploration

Xingxing Xie, Heliya Ahmadi, Dylan Paré
xkx5097@psu.edu, hxa5384@psu.edu, dpare@psu.edu
The Pennsylvania State University

Abstract: This case study explores how interactional tools can support the development of design ideas within a collaborative virtual reality (VR)-based studio. Six participants explored multiplayer VR applications to reflect on existing VR designs and co-design an LGBTQ+ cultural mini-golf game, while documenting their experiences in reflective journals. Our findings extend sociocultural constructionist and socio-materialist perspectives, demonstrating how VR pens, viewed as "non-neutral actors," shape learners' design idea generation, refinement, and decision-making.

Introduction

Virtual reality (VR) enables users to step into computer-generated, three-dimensional environments where they can interact with objects and experience a strong sense of "presence" (Mayer et al., 2023). Previous work has demonstrated the unique affordances of VR for learning, such as supporting experiential learning tasks, making abstract concepts tangible, and enhancing learners' spatial knowledge (e.g., Silseth et al., 2024). Moreover, VR enables learners to use their bodies and gestures to manipulate objects within 3D spaces, which supports embodied learning and enhances their sense of agency over the learning process (Johnson-Glenberg, 2018, 2019).

Researchers have shown growing interest in how VR can support collaborative learning across disciplines (Fortman & Quintana, 2023; Paulsen et al., 2024). Paré et al. (2019), for example, demonstrated how creating 3D sculptures and customizing avatars in multiplayer VR settings supported playful inquiry, enabling participants to collaboratively explore and share queer experiences. Yet, immersive VR is still often perceived as an individual technology, with many learning environments designed for the individual (e.g., Enyedy & Yoon, 2021). Much of the existing research also focuses on evaluating learning outcomes, rather than on the learning processes themselves (Radianti et al., 2020). Few studies employ qualitative methods to examine how learning occurs in collaborative VR, and how participants utilize embodied and interactional resources to co-construct meaning (Paulsen et al., 2024). Furthermore, most research focuses on learners engaging with pre-designed VR content to acquire specific procedural or declarative knowledge (Radianti et al., 2020), with far less attention paid to learners' inventive and design-oriented practices as creators in collaborative VR environments.

Theoretical background and objectives

From a constructionist perspective, learning occurs when learners actively engage in constructing a "public entity" that can be shared with others. We thus conceptualize the objects learners created in the VR worlds as not merely knowledge application products, but as powerful "objects-to-think-with" (Papert, 1980) created through embodied VR learning experiences that support their decision-making. We also draw upon socio-materialism (Barad, 2003), which argues for a shift in the transitional view from "how humans use tools" to "how humans and materials act together" (Peppler & Thompson, 2024). It challenges the traditional view of tools and materials as passive intermediaries, which posits that tools and materials are "non-neutral actors" possessing their agency, historicity, and identity (Peppler & Thompson, 2024). That is, these tools are not just passively utilized by human beings, but rather proactively co-shape learning and creative processes alongside them. This process aligns with Barad's (2003) notion of "intra-action," emphasizing that humans and tools co-produce meaning and outcomes through mutual interaction, rather than one unilaterally acting upon the other. We focus on a specific design project to understand how individuals learn about VR and cultural heritage design more broadly through participation in a collaborative VR studio. We then ask: *In collaborative VR exploration, how do interaction tools support individual idea development when designing a VR cultural mini-golf game?*

Methods

We employed a qualitative case study (Yin, 2018), situated within a larger participatory design research project. The case focuses on a six-week VR studio as a particular, bounded design context where six participants collaboratively explored VR design with the goal of designing an LGBTQ+ historical and cultural mini-golf game. Participants included three graduate students and three VR industry professionals from Canada and the United

 © ISLS

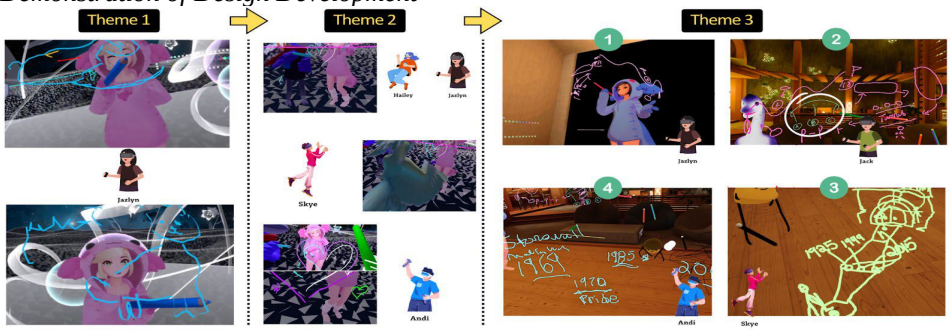
States. They utilized VR headsets to explore multiplayer applications (e.g., VRChat) and engage in design prototyping (e.g., ShapesXR). We specifically examined the use of VR pens as interactional tools. Unlike 2D drawing, VR pens allow for dynamic computational geometry creation in 3D space, heavily relying on learners' physical movements and gestures. The data comprised forty-four reflective journals, eight video recordings, seventy-one participant photographs, and thirty-three-dimensional artifacts created within the virtual environment. For this analysis, the individual's interactional tool-use episode and its support of collaborative design serve as the primary unit of analysis. We then followed Braun and Clarke's (2006) thematic analysis, coding specifically for how interactional tool affordances shaped design development.

Findings

We found three main themes indicating that VR pens function as "non-neutral actors" that explicates how VR pens supported participants' design idea development: (1) Making the "invisible" visible; (2) Sketches as "Prisms"; and (3) "Anchors" as Design Stages. We present one of the participants, Jazlyn, to illustrate these themes, as it provides a rich vignette showcasing how pens facilitated her design thinking across different design stages, from individual to collaborative processes. First, the tools made participants' thinking visible by externalizing intangible ideas into concrete 3D sketches. For example, Jazlyn used the VR pens to materialize personal memories into visible artifacts (see Figure 1, left-hand column), transforming internal thoughts into shareable "objects-to-think-with" (Papert, 1980) as entry points for subsequent game design generation.

Figure 1

Demonstration of Design Development



Second, sketches functioned as "prisms" rather than mirrors of thought, refracting design ideas into multiple perspectives that were not apparent in verbal discussions alone. In Jazlyn's reflective journal discussing her experience, for example, the act of her drawings unfolded into different lenses (e.g., personal story sharing, co-creation, and freedom of expression), refining her understanding of how ideas could be represented in VR space. Through a socio-material lens, these refractions emerged from the intra-action between learners and the tool, where the virtual material properties of the pen co-produced new design dimensions (Barad, 2003). Third, sketches served as potential "anchors" for group discussions and design alignment. Jazlyn's sketches of the mini-golf game design (see Figure 1, right-hand column, number 1) served as reference points that support shared, iterative design stages, enabling the group to converge on collective decisions. Through Jack's integration of game mechanics, Andi's historical narratives, and Skye's refined sketch, the artifact evolved from a personal expression into a "public entity" (Papert, 1980). This showcases a practice of VR co-creation, where the modifiability of VR pens allowed for the visible integration of diverse contributions within the shared space.

Conclusion

Our case study shows VR interactional tools within collaborative VR environments actively shape the trajectory of design development rather than serving as passive instruments. By enabling learners to externalize invisible concepts, refract ideas through multiple perspectives, and anchor collective iterations, VR pens function as "non-neutral actors" that fundamentally alter the design process. From constructionist and socio-material lenses, our findings demonstrate that the intra-action between learners and the specific affordances of VR pens, such as 3D spatial freedom and lack of physical resistance, co-produces unique forms of "reflection-in-action" (Schön, 1983) and collaborative knowledge building. We hope these insights will inspire researchers, practitioners, and designers in the field of learning sciences to rethink the role of tools' materiality and non-neutrality in empowering and shaping more creative, diverse, and inclusive educational environments for the future.

 © ISLS

References

Barad, K. (2003). Posthumanist performativity: Toward an understanding of how matter comes to matter. *Signs: Journal of Women in Culture and Society*.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101.

Enyedy, N., & Yoon, S. (2021). Immersive environments: Learning in augmented + virtual reality. In *International Handbook of Computer-Supported Collaborative Learning* (pp. 389–405). Springer.

Fortman, J., & Quintana, R. (2023). Fostering collaborative and embodied learning with extended reality: Special issue introduction. *International Journal of Computer-Supported Collaborative Learning, 18*(2), 145–152.

Johnson-Glenberg, M. C. (2018). Immersive VR and education: Embodied design principles that include gesture and hand controls. *Frontiers in Robotics and AI*, 5.

Johnson-Glenberg, M. C. (2019). The necessary nine: Design principles for embodied VR and active STEM education. In *Learning in a Digital World* (pp. 83–112). Springer.

Mayer, R. E., Makransky, G., & Parong, J. (2023). The promise and pitfalls of learning in immersive virtual reality. *International Journal of Human–Computer Interaction, 39*(11), 2229-2238.

Papert, S. (1980). *Mindstorms: Children, computers, and powerful ideas*. Basic Books.

Paulsen, L., Dau, S., & Davidsen, J. (2024). Designing for collaborative learning in immersive virtual reality: A systematic literature review. *Virtual Reality, 28*(1), 1–17.

Paré, D., Sengupta, P., Windsor, S., Craig, J., & Thompson, M. (2019). Queering virtual reality: A prolegomenon. In P. Sengupta, M.-C. Shanahan, & B. Kim (Eds.), *Critical, transdisciplinary and embodied approaches in STEM education* (pp. 307–328). Springer.

Peppler, K., & Thompson, N. (2024). Tools and materials as non-neutral actors in STEAM education. *Journal of the Learning Sciences, 33*(4–5), 719–756.

Radianti, J., Majchrzak, T. A., Fromm, J., & Wohlgenannt, I. (2020). A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers & Education*, 147, 103778.

Schön, D. A. (2017). *The reflective practitioner: How professionals think in action*. Routledge.

Silseth, K., Steier, R., & Arnseth, H. C. (2024). Exploring students' immersive VR experiences as resources for collaborative meaning making and learning. *International Journal of Computer-Supported Collaborative Learning, 19*(1), 11–36.

Yin, R. K. (2011). *Applications of case study research* (3rd ed.). Sage Publications.

International Society of the Learning Sciences

At the Intersection of Cognitive Engagement and Critical Thinking in AI-Assisted Learning

Lu Ding, Erika Cheek, Tami May
luding@southalabama.edu, echeek@southalabama.edu, tmay@southalabama.edu
Counseling and Instructional Sciences, University of South Alabama, USA

Abstract: This study examined how cognitive engagement levels, defined by the ICAP framework, relate to critical thinking in student-AI interactions. Analyzing 784 student questions submitted to a GPT-4-powered chatbot in an introductory biology course, we coded each question for ICAP level and applied Newman et al.'s (1995) critical-thinking indicators. ANOVA results revealed significant differences across ICAP categories: Constructive questions yielded the highest critical-thinking ratios, followed by Active and Passive. However, Constructive questions were mostly clarifying or confirmatory, rarely exhibiting evaluative reasoning. These findings suggest that higher ICAP engagement predicts higher critical thinking, but students often remain at explanatory rather than evaluative levels when interacting with LLMs. The findings shed light on the need for instructional scaffolds that promote both cognitive engagement and epistemic vigilance.

Introduction

As generative AI tools become increasingly integrated into education, researchers emphasize the need to move beyond answer-seeking toward deeper learning and critical thinking. LLMs can provide instant, tailored responses, but their educational value depends heavily on how students engage with them (Jošt et al., 2024). A key concern is that students may over-rely on AI-generated information and accept responses uncritically, hindering higher-order thinking (Darvishi et al., 2024). In LLM-based learning, students—not instructors—drive the questioning process, making question depth a central determinant of outcomes.

In our prior work (Ding et al., 2026), student questions to ChatGPT were coded using ICAP, finding that higher-performing students asked significantly more Constructive questions while lower-performers asked more Passive ones. The present study extends this by examining whether different ICAP question types also correspond to differences in critical-thinking quality, using Newman et al.'s (1995) framework. We address two research questions: (1) How does question type relate to critical thinking quality in student-AI learning interactions? (2) What patterns of critical thinking indicators emerge within Constructive questions?

Literature review

Wittrock's (1992) Generative Learning Theory (GLT) holds that learning depends on what learners mentally construct rather than passively receive. Building on GLT, Chi and Wylie's (2014) ICAP framework classifies cognitive engagement across four modes, Interactive, Constructive, Active, and Passive, each representing an increasing degree of generative processing, with empirical support showing that higher-mode engagement produces stronger learning outcomes (Menekse et al., 2013; Morris & Chi, 2019). Student questioning is a key driver of this engagement. Explanation-seeking questions reflect higher-order reasoning, while factual questions signal surface-level processing (Graesser et al., 2003), mapping directly onto ICAP's Constructive/Interactive and Passive/Active modes, respectively. Structured scaffolds are needed to shift students toward deeper questioning (King, 1994; Chin & Osborne, 2008), and this is equally true in AI-mediated contexts, as Cheng et al. (2025) found that prompt quality significantly shapes student performance with GenAI. Critical thinking is essential in these contexts precisely because LLM outputs, though fluent, may be inaccurate or biased (Cain, 2024). Learners must practice epistemic vigilance, assessing the reliability and coherence of AI-generated information rather than accepting it uncritically (Sperber et al., 2010; Bielik & Krell, 2025). Calibrated trust is the goal: neither overtrust that yields passive acceptance of misinformation, nor excessive skepticism that forecloses productive AI use.

Methodology

This study was conducted in an introductory biology course with an enrollment of 140 students at a southeastern U.S. university, of which 85 voluntarily participated in the study (58.8% female, 39.2% male, 2.0% non-binary; 46.1% White, 24.5% Black, 7.8% Hispanic, 4.9% Asian). After the first exam in week 5 of the 16-week course, students discussed their incorrect answers with a custom GPT-4o biology tutor chatbot, and a total of 784 pairs of interactions were generated.

 © ISLS

Each of the student questions was coded using an ICAP rubric (Ding et al., 2026a) as Interactive, Constructive, Active, or Passive, then independently coded using Newman et al.'s (1995) critical thinking framework. The framework consists of ten positive indicators (e.g., relevance, justification) and ten negative indicators (e.g., repetition, uncritical agreement), which can yield a CTRatio where higher values indicate greater critical reasoning. Two trained research assistants (RAs) first independently coded a randomly selected 10% of the questions and met with the first author to discuss any discrepancies and then coded another set of 10% randomly selected questions and reached a high consistency in their coding. The remaining 80 percent of the data was divided in half, and each RA coded one half independently. Due to the fact that only three Interactive questions were asked by students, they were excluded from the analysis. ANOVA with post hoc LSD comparisons examined CTRatio differences across ICAP categories, followed by a deeper indicator-level analysis of Constructive questions.

Results

ANOVA revealed a significant main effect of question type on CTRatio, $F(2, 781) = 32.70$, $p < .001$, partial $\eta^2 = .14$. Constructive questions had the highest mean score ($M = .57$, $SD = .54$, $n = 129$), followed by Active ($M = .34$, $SD = .56$, $n = 279$) and Passive ($M = .07$, $SD = .39$, $n = 373$). Post hoc LSD comparisons confirmed all group differences were statistically significant ($p < .001$).

Deeper analysis of the 129 Constructive questions (387 total codes) revealed that most were characterized by relevance (R+) and clarification-seeking (A+), indicating students stayed on topic and sought to resolve ambiguities. About half introduced new problem-related information (NP+), yet nearly one-third contained judgments without justification (JS−), proof/examples (J+) appeared in only 0.8% of questions, and critical assessment (C±) was never coded, suggesting students engaged generatively but rarely reached evaluative reasoning (see Table 1).

Table 1

Top Five and Last Five Used Codes In C-Type of Questions

Code Meaning	Freq.	Perc.	Code Meaning	Freq.	Perc.
Relevant statements	122	94.6%	Relate solutions to familiar situations	1	0.8%
Clear up ambiguities	84	65.1%	Clear, unambiguous statements	1	0.8%
New problem-related information	65	50.4%	Providing proof/examples	1	0.8%
Judgments without justification	37	28.7%	Learner brings new things in	1	0.8%
Linking ideas, making inferences	30	23.3%	Using previous knowledge	1	0.8%

Discussion and implication

The results corroborate ICAP premises, showing that critical thinking depth increased systematically from Passive to Active to Constructive questioning. Constructive prompts produced the highest mean critical-thinking scores, indicating that higher cognitive engagement was associated with deeper reasoning. However, the mean CTRatio for C questions ($M = .57$) is still modest. Most Constructive questions were relevant (R+) and clarifying (A+), demonstrating that students were actively verifying understanding and addressing ambiguities. Yet nearly one-third exhibited judgments without justification (JS−), providing proof/examples (J+) was one of the least used codes, and critical assessment (C±) never appeared, suggesting a tendency toward reasoning-seeking rather than reasoning-producing discourse. These patterns align with Graesser et al. (2003) that learners often use explanation-seeking questions to confirm knowledge rather than generate new inferences. Our results suggest that even when students constructively engage with the LLM, their engagement does not always reach higher levels of critical thinking.

This finding implies that Constructive questioning may plateau at an explanatory level rather than extending toward evaluative reasoning or critique. This could be because students treat GenAI as an authoritative source (Ding et al., 2026b), making their questioning confirmatory rather than scrutinizing. Given the hallucination effect of LLMs, it is dangerous to interact with LLMs without epistemic vigilance. These dynamics underscore the dual challenge of fostering cognitive engagement while cultivating skepticism toward AI-generated explanations, highlighting the need for future work to examine how to design instructional training that will **engage students cognitively** with LLMs for their learning and foster students' **epistemic vigilance**. Future research should conduct intervention studies to test whether explicitly teaching questioning strategies that emphasize providing proof or examples, making justified judgments, and relating solutions to familiar contexts can cultivate epistemic vigilance while fostering deeper engagement in AI-mediated learning.

References

Bielik, T., & Krell, M. (2025). Developing and evaluating the extended epistemic vigilance framework. Journal of Research in Science Teaching, 62(3), 869–895. https://doi.org/10.1002/tea.22100

Cain, W. (2024). Prompting change: Exploring prompt engineering in large language model AI and its potential to transform education. TechTrends, 68(1), 47–57. https://doi.org/10.1007/s11528-023-00896-0

Cheng, Y., Lai, C.-F., Chen, Y.-T., Wang, W.-S., Huang, Y.-M., & Wu, T.-T. (2025). Asking generative artificial intelligence the right questions improves writing performance. Computers and Education: Artificial Intelligence, 8, 100374. https://doi.org/10.1016/j.caeai.2025.100374

Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. Educational Psychologist, 49(4), 219–243. https://doi.org/10.1080/00461520.2014.965823

Chin, C., & Osborne, J. (2008). Students' questions: A potential resource for teaching and learning science. Studies in Science Education, 44(1), 1–39.

Darvishi, A., Khosravi, H., Sadiq, S., Gašević, D., & Siemens, G. (2024). Impact of AI assistance on student agency. Computers and Education, 210, 104967. https://doi.org/10.1016/j.compedu.2023.104967

Ding, L., O'Berry, R., Tallent, H., Chhetri GC, S. G., & Gapud, A. (2026b). Learning with large language models: Beyond prompt engineering. Education and Information Technologies. https://doi.org/10.1007/s10639-026-13924-2

Ding, L., Rodenberg, R., Er, E., & Nguyen, H. (2026a). *Learning through generative questioning with generative AI: An ICAP approach*. Paper presented at the annual meeting of the American Educational Research Association, Los Angeles, CA, USA.

Graesser, A. C., & Olde, B. A. (2003). How does one know whether a person understands a device? The quality of the questions the person asks when the device breaks down. Journal of Educational Psychology, 95(3), 524–536. https://doi.org/10.1037/0022-0663.95.3.524

Jošt, G., Taneski, V., & Karakatič, S. (2024). The impact of large language models on programming education and student learning outcomes. Applied Sciences, 14(10), 4115. https://doi.org/10.3390/app14104115

King, A. (1994). Guiding knowledge construction in the classroom: Effects of teaching children how to question and explain. American Educational Research Journal, 31(2), 338–368.

Menekse, M., Stump, G. S., Krause, S., & Chi, M. T. H. (2013). Differentiated overt learning activities for effective instruction in engineering classrooms. Journal of Engineering Education, 102(3), 346–374. https://doi.org/10.1002/jee.20021

Morris, B. J., & Chi, M. T. H. (2019). Training teachers to move from active to constructive and interactive modes of instruction. Educational Psychologist, 54(3), 219–242.

Newman, D. R., Webb, B., & Cochrane, C. (1995). A content analysis method to measure critical thinking in face-to-face and computer-supported group learning. Interpersonal Computing and Technology, 3(2), 56–77.

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). Epistemic vigilance. Mind & Language, 25(4), 359–393.

Wittrock, M. C. (1992). Generative learning processes of the brain. Educational Psychologist, 27(4), 531–541. https://doi.org/10.1207/s15326985ep2704_8

International Society of the Learning Sciences

ClassMosaic: Leveraging Solution Diversity to Teach Computational Problem-Solving

Laila El-Hamamsy, University of Teacher Education (HEP) Vaud, Switzerland, laila.elhamamsy@hepl.ch
Engin Bumbacher, University of Teacher Education (HEP) Vaud, Switzerland, engin.bumbacher@hepl.ch

Abstract: A key challenge in K-12 Computer Science (CS) education is teaching computational Problem-Solving (PS). PS instruction research emphasizes the importance of structured whole-class discussions to reflect on solution strategies to foster PS competences. However, orchestrating whole-class discussions is challenging for teachers. As such, we drew on PS- and dashboard research to design ClassMosaic, a classroom orchestration tool that supports productive PS-teaching practices. ClassMosaic provides real-time insights into student work to facilitate discussions and promote collaborative learning by enabling collective comparison and reflection. We present ClassMosaic's design principles and the first user study focusing on perceived pedagogical value with 9 CS teachers. Results show that ClassMosaic aligns with existing CPS instruction practices, and fosters new CPS-teaching strategies, including heterogeneous strategy-based grouping and structured post-discussion adaptations. Future work will further automate PS-based code clustering, and adapt ClassMosaic's UX/UI to reduce teacher cognitive load, before in-class experiments.

Introduction and related work

PS is a core competence in CS that students struggle to acquire (Medeiros et al., 2019; Pechorina et al., 2023). As research on explicitly teaching Computational PS (CPS) has produced mixed results (Craig et al., 2019; Denny et al., 2019; Pechorina et al., 2023; Prather et al., 2020), how to design and structure CPS teaching to foster learning and transfer of effective CPS competences remains an open question. While research on teaching PS in K-12 CS education is still emerging, mathematics education can provide relevant insight due to its decades-long research into teaching mathematical PS (Santos-Trigo, 2024; Silver, 2023; Stanic and Kilpatrick, 1989).

A core finding from mathematics education is the importance of (1) having open-ended, cognitively engaging PS activities with multiple solutions (Lester and Cai, 2016), and (2) orchestrating whole-class discussions (Stein et al., 2008) to reflect on and compare the diversity of solutions. Whole-class discussions contribute significantly to developing students' PS competences (Silseth and Furberg, 2024; Stein et al., 2008) through (i) sense-making and reflection, (ii) consolidating knowledge, and (iii) deepening students' understanding (Ackermann, 2001). However, orchestrating whole-class discussions is challenging for teachers (Evans and Dawson, 2017; Liljedahl, 2016). This is where teaching CPS provides an interesting opportunity: given the central role of digital technologies in CS, it becomes relevant to study how technological tools enable and support effective PS teaching practices (Quintana et al., 2004; Santos-Trigo and Moreno-Armella, 2016).

Despite increasing research on dashboards for education, there are frequently raised limitations: (1) few dashboards are designed and adapted to K-12 pedagogical practices, (2) lack of theoretical grounding (Hundhausen et al., 2017; Jivet et al., 2017), and (3) focus on awareness (Jivet et al., 2017) and / or prediction, without a clear translation between the analytics and pedagogical actions, thus limiting "the educational value" Masiello et al., 2024, p.8 of dashboards. To address these limitations we developed ClassMosaic.

ClassMosaic

ClassMosaic is a web-based platform designed to support teachers in CPS instruction by facilitating productive whole-class CPS discussions and leveraging the diversity of students' solutions to support collective sense-making and learning. To that effect, ClassMosaic makes students' PS strategies and solutions visible, comparable, and interpretable in real time. ClassMosaic's design draws from (i) mathematics PS research to provide a theoretical foundation, and (ii) dashboard research to adhere to best practices and address challenges that affect teacher use and adoption. The goal is to enable teachers to engage with student data to support pedagogical decision-making in real-time, and thus extend the role of dashboards in CPS education beyond awareness tools.

ClassMosaic is compatible with Scratch (Resnick et al., 2009, development underway for python) and comprises of three main parts: (1) The task designer to create single or multi-task programming sessions where teachers can (a) limit the set and number of available blocks, (b) freeze selected code blocks, and (c) create custom code tests (assertions) to evaluate students' solutions. (2) The automatic real-time collection of students' codes in-development. (3) The interactive dashboard, which can be viewed by the teacher alone or together with the classroom, enables real-time engagement with student data from the CPS activity in multiple ways: (a) An

 © ISLS

interactive 2D graph view shows student codes arranged according to different code metrics to make visible the commonalities and rich variations in solution approaches. Teachers can assign codes manually to groups, or choose automated groupings of student code (based on clustering of Abstract Syntax Trees). (b) A code comparison view that shows manually or automatically selected codes side-by-side. These elements aim to support real-time exploration, identification, and analysis of students' CPS approaches.

Materials and methods

We conducted a preliminary user study with 9 K-12 and higher education CS teachers, all with extensive teacher training experience and ongoing CS teaching mandates, to examine ClassMosaic's perceived pedagogical value. The user-study was organized as follows: (1) Questions regarding pre-existing CS and CPS teaching practices and needs; (2) Solving a pre-designed Scratch task with ClassMosaic; (3) Using ClassMosaic to explore a pre populated solution data set for the same task while following a think-aloud protocol; (4) Feedback on the user experience to identify emerging use cases, affordances and barriers; (5) Synthesis of key-findings for participant validation to minimize misinterpretations (Merriam and Tisdell, 2015).

The data analysis was conducted by two researchers as follows (Saldaña, 2009): (1) A descriptive coding phase to classify the topics in each passage. (2) A focused coding phase to create more refined codes within each topic in terms of teacher practices, and features needed to support them. (3) Grouping teacher practices into high-level categories (lesson preparation, classroom & data management, CPS instruction, analysis and feedback).

Results and discussion

Existing practices & needs to teach CPS. Prior to seeing ClassMosaic, the teachers already engaged in some productive CPS instructional practices, but needed technological support, including (i) a system where students can solve the same problem in different programming languages, reinforcing CPS beyond language syntax; (ii) awareness of classroom activity through direct access to students' codes and progress; (iii) automated grading.

ClassMosaic in support of existing practices. ClassMosaic was perceived as facilitating classroom management and orchestration by making student data accessible, easy to display, compare, execute, and edit. Teachers noted that this would enable them to (1) view student work in real-time without interrupting them, and (2) identify and showcase examples in post-task reflections. The code analysis and grouping features were seen as helpful to identify strategies and enrich whole-class discussions. However, teachers requested more automated grouping functionalities to streamline their workflow and minimize in-class time spent behind the screen.

ClassMosaic in support of novel practices. Beyond enhancing existing practices, ClassMosaic prompted teachers to consider new instructional approaches, particularly to analyse student work and adapt lesson strategies. Notably, after interacting with ClassMosaic more teachers explicitly mentioned the relevance of getting students to compare and reflect on the differences between them. Teachers also reflected on ClassMosaic's potential to support CPS instruction by (1) creating heterogeneous groups based on PS strategies; (2) adjusting subsequent lesson designs using data insights; (3) providing individualized formative and summative feedback.

By supporting established and emerging CPS teaching practices, ClassMosaic helps bridge the gap between theoretical best practices and real-world classroom needs. However, refinements are required to improve ClassMosaic's potential utility and effectiveness to teach CPS by: (1) reducing teachers' in-class cognitive load to analyse students' strategies (e.g., fully automated code clustering); (2) preparing teachers for in-class use of ClassMosaic to ensure that the practices that teachers engage in align with pedagogical best practices that have been found to foster PS competences (Leeuwen et al., 2021; Molenaar and Knoop-van Campen, 2019).

Limitations As for all studies, the current study has its limitations: (1) small sample of participants with a deep understanding of best practices and novel teaching tools; (2) focus on a single programming language (Scratch); (3) perception based on limited use as teachers' feedback was based on unguided interactions with pre-collected data, not actual classroom use, which hindered their ability to anticipate student responses, increased the need for guidance, and reduced the effectiveness of data exploration with the tool.

Conclusion and future work

Novel technological approaches to CPS instruction with theory-informed design are perceived by teachers to facilitate existing teaching practices and enable new instructional strategies. Teachers considered that ClassMosaic could be valuable to support classroom orchestration, whole-class discussions, and comparative analyses of students' solutions. Teachers also identified features that could enhance ClassMosaic's utility, such as automated solution grouping, cross-language comparisons, and integrated formative feedback mechanisms. The variety of teaching practices evoked also reflects the value of designing for pedagogical flexibility, so a tool can be more effectively and expand on different teaching practices. These results reinforce the importance of aligning educational technology with pedagogical best practices and involving teachers in the design process.

References

Ackermann, E. (2001). Piaget's constructivism, Papert's constructionism: What's the difference. Future of learning group publication, 5(3), 438.

Craig, M., Petersen, A., & Campbell, J. (2019). Answering the Correct Question. CompEd'19, 72–77.

Denny, P., Prather, J., Becker, B. A., Albrecht, Z., Loksa, D., & Pettit, R. (2019). A Closer Look at Metacognitive Scaffolding: Solving Test Cases Before Programming. Koli Calling'19, 1–10.

Evans, S., & Dawson, C. (2017). Orchestrating Productive Whole Class Discussions: The Role of Designed Student Responses. Mathematics Teacher Education and Development, 19(2), 159–179.

Hundhausen, C. D., Olivares, D. M., & Carter, A. S. (2017). IDE-Based Learning Analytics for Computing Education: A Process Model, Critical Review, and Research Agenda. ACM Trans. Comput. Educ.

Jivet, I., Scheffel, M., Drachsler, H., & Specht, M. (2017). Awareness Is Not Enough: Pitfalls of Learning Analytics Dashboards in the Educational Practice. Data Driven Approaches in Digital Education, 82–96.

Leeuwen, A. v., Campen, C. A. N. K. -v., Molenaar, I., & Rummel, N. (2021). How Teacher Characteristics Relate to How Teachers Use Dashboards: Results From Two Case Studies in K-12. Journal of Learning Analytics.

Lester, F. K., & Cai, J. (2016). Can Mathematical Problem Solving Be Taught? Preliminary Answers from 30 Years of Research. In Posing and Solving Mathematical Problems: Advances and New Perspectives (pp. 117–135). Springer.

Liljedahl, P. (2016). Building Thinking Classrooms: Conditions for Problem-Solving. In Posing and Solving Mathematical Problems: Advances and New Perspectives (pp. 361–386). Springer.

Masiello, I., Mohseni, Z. (, Palma, F., Nordmark, S., Augustsson, H., & Rundquist, R. (2024). A Current Overview of the Use of Learning Analytics Dashboards. Education Sciences, 14(1), 82.

Medeiros, R. P., Ramalho, G. L., & Falc ˜ao, T. P. (2019). A Systematic Literature Review on Teaching and Learning Introductory Programming in Higher Education. IEEE Trans. Educ., 62(2), 77–90.

Merriam, S. B., & Tisdell, E. J. (2015). Qualitative research: A guide to design and implementation (4th). John Wiley & Sons.

Molenaar, I., & Knoop-van Campen, C. A. N. (2019). How Teachers Make Dashboard Information Actionable. IEEE T LEARN TECHNOL, 12(3).

Pechorina, Y., Anderson, K., & Denny, P. (2023). Metacodenition: Scaffolding the Problem-Solving Process for Novice Programmers. ACE'23.

Prather, J., Becker, B. A., Craig, M., Denny, P., Loksa, D., & Margulieux, L. (2020). What do we think we think we are doing? metacognition and self-regulation in programming. ICER'20, 2–13.

Quintana, C., Reiser, B. J., Davis, E. A., Krajcik, J., Fretz, E., Duncan, R. G., Kyza, E., Edelson, D., & Soloway, E. (2004). A Scaffolding Design Framework for Software to Support Science Inquiry. In Scaffolding. Psychology Press.

Resnick, M., Maloney, J., Monroy-Hernández, A., Rusk, N., Eastmond, E., Brennan, K., Millner, A., Rosenbaum, E., Silver, J., Silverman, B., et al. (2009). Scratch: Programming for all. Communications of the ACM.

Saldaña, J. (2009). The coding manual for qualitative researchers. Sage.

Santos-Trigo, M. (2024). Problem solving in mathematics education: Tracing its foundations and current research-practice trends. ZDM Mathematics Education, 56(2), 211–222.

Santos-Trigo, M., & Moreno-Armella, L. (2016). The Use of Digital Technology to Frame and Foster Learners' Problem-Solving Experiences. In Posing and Solving Mathematical Problems: Advances and New Perspectives. Springer.

Silseth, K., & Furberg, A. (2024). Bridging group work and whole-class activities through responsive teaching in science education. European Journal of Psychology of Education, 39(3), 2155–2176.

Silver, E. A. (2023). Conclusion: Mathematics Problem Posing and Problem Solving: Some Reflections on Recent Advances and New Opportunities. In Problem Posing and Problem Solving in Mathematics Education: International Research and Practice Trends (pp. 247–259). Springer Nature.

Stanic, G., & Kilpatrick, J. (1989). Historical perspectives on problem solving in the mathematics curriculum. The teaching and assessing of mathematical problem solving, 3, 1–22.

Stein, M. K., Engle, R. A., Smith, M. S., & Hughes, E. K. (2008). Orchestrating Productive Mathematical Discussions: Five Practices for Helping Teachers Move Beyond Show and Tell. Math Think Learn.

International Society of the Learning Sciences

Multimodal Analysis of Mathematical Statistics Learning in a CSCL Environment

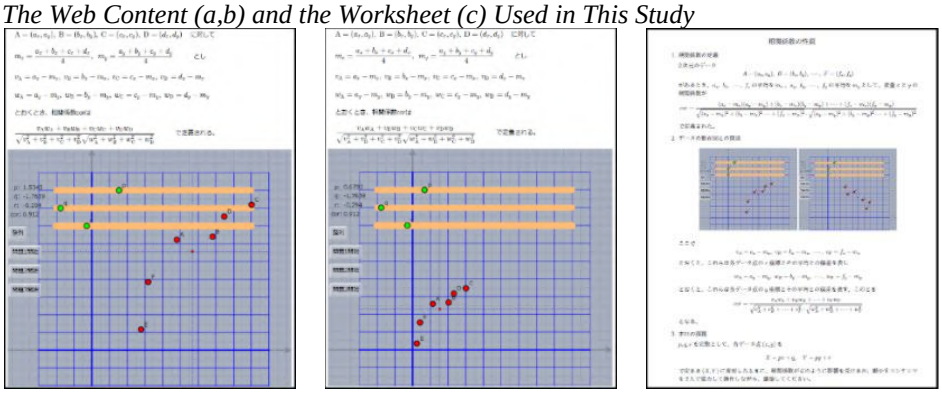
Masataka Kaneko, Toho University, Masataka.kaneko@phar.toho-u.ac.jp
Hironori Egi, The University of Electro-Communications, hiro.egi@uec.ac.jp
Takuya Kitamoto, Yamaguchi University, kitamoto@yamaguchi-u.ac.jp
Takeo Noda, Toho University, noda@c.sci.toho-u.ac.jp

Abstract: This study, a multimodal analysis of learning data collected using a video recorder, high-resolution microphone, and dynamic geometry system, focused on pairs of university students engaged in a mathematical statistics task while collaboratively using HTML-based dynamic content. The qualitative analysis of transcribed text data from participants' discourse was based on Sfard's commognitive perspective. The results were then compared to the coded data of their behavior, the log data of their manipulating the dynamic content, and the audio data of their conversations. Our findings indicate that there is a relationship between the transition of learners' behavioral patterns and the emergence of refined discursive routines.

Introduction and research method

While mathematical thinking is grounded in the interplay between cerebral networks for language processing and those for visuospatial processing (Dehaene et al., 1999), it emerges also from the mental simulation of dynamic imagery, which is supported by the interlocking network of the visual system and the motor system (Lakoff & Nunez, 2001). Therefore, Multimodal Learning Analytics (MMLA) should be an indispensable approach to investigate the CSCL process in math. As indicated in a previous study (Schwarz et al., 2018), there might exist some "critical moments" which are identified through the occurrence of certain types of verbal behaviors. Accordingly, this study investigates the verbal and nonverbal behaviors of several pairs of university students as they were engaged in a statistical task to uncover how the qualitative change in participants' discourse is projected onto the temporal transition of their observed behavior. To elucidate the progress of learners' thinking, we employed commognitive perspective (Sfard, 2008). Collaboratively using an HTML-based dynamic content in Figure 1 (left and center), created by using the dynamic geometry system CindyJS (https://www.cindyjs.org), they were asked to consider how the correlation coefficient associated with the data points $p_i(x_i,\ y_i)$ $(i = 1,2,3,\dots,n)$ changes under the affine transformation $X_i = px_i + q,\ Y_i = py_i + r$. In this topic, discursive routine including "spread of data points" to represent the correlation coefficient can lead to commognitive conflict. When the data points and the coefficients p, q, r are manipulated, the resulting points $(X_i,\ Y_i)$ and correlation coefficient are displayed instantly. The manipulation process is logged with time stamps. Figure 1 (right) shows the worksheet in which related mathematical expressions are included and the participants made a note. Also, the audio data is used to plot the temporal transition of the quantity of each member's utterances on a timeline by using the dialogue analysis system (Ishikawa et al., 2019). Moreover, participants' behaviors identified in the video recordings were classified into several categories, including "manipulation of dynamic content", "pointing to the symbolic representations", "making gestures", and "writing on the worksheet". Classified behaviors of each participant were coded on a timeline by using the Sportscode system (https://www.hudl.com/products/sportscode).

Figure 1
The Web Content (a,b) and the Worksheet (c) Used in This Study



(a) (b) (c)

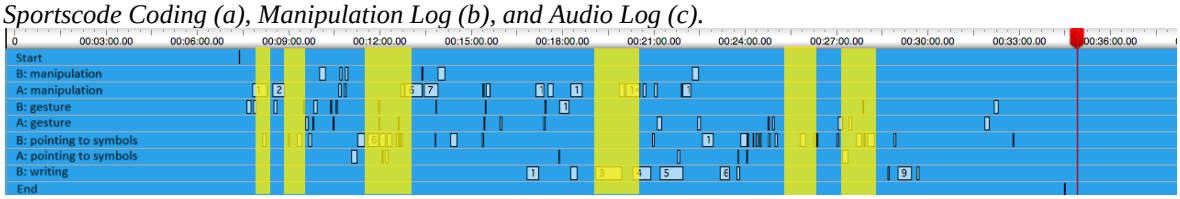
 © ISLS

Result and discussion

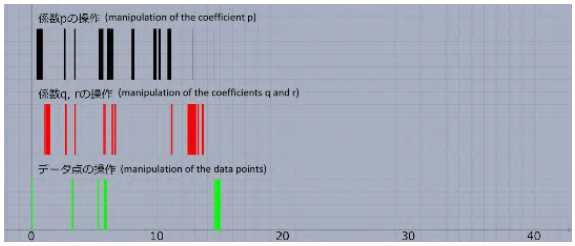
Figure 2 shows the visualization of timelines for a group. The time intervals with the progress in thinking, which are identified by analyzing the transcribed text data, are specified by using yellow in the Sportscode timeline. In the visualization of manipulation log, the timings when the first-order coefficient p, zeroth-order coefficients q, r, and data points are manipulated are specified by using black, red, and green, respectively. In the visualization of audio log, while the temporal transition in the quantity of each participant's utterances is plotted by using column charts, change in the power balance between participants is specified by the orange curve. Also, the time intervals with the progress in thinking are specified by using light green color.

Figure 2

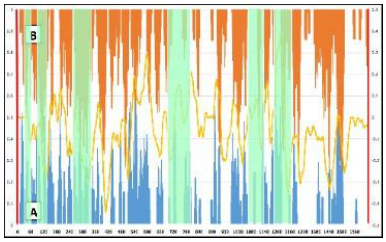
Sportscode Coding (a), Manipulation Log (b), and Audio Log (c).



(a)



(b)



(c)

During the first two intervals, participants mainly compared the symbolic expression and dynamic content with utterances for target setting. In contrast, in the third interval, participants' behavior moved between the manipulation of dynamic content and pointing to symbols with a new discursive routine involving symbolic expressions and the word "mean value". As seen in lower right, the quantity of both participants' utterances remarkably increased along with this change. Contrarily, in the former half of the fourth interval, a new discursive routine using the word "the ratio of the distance" related to the geometric configuration emerged while manipulation by participant A and writing by participant B continued. Then, in the latter half, this routine led to more refined one for certifying the effect of manipulating data points, such as "The distance is maintained since the average moves in the same amount" with the increase of utterance from A. Next, in the fifth interval, the discussion shifted to a more refined routine, including the phrases as "square root of the quantity multiplied by p^2", "the factor p in both denominator and numerator can be cancelled". A small amount of behavior and utterance indicates that the thinking remained in a germinal stage. In fact, in the last interval, a conclusive routine including "standard deviation in the denominator", "covariance in the numerator", and "reduction of fraction with the same factors", together with pointing to the symbolic representations, appeared while the dominance of utterance largely shifted from A to B. Comparison with the data from other groups shows that the emergence of reified discursive routines refrained participants from manipulating data points at an earlier stage as seen in lower left.

Thus, the emergence of refined discursive routines is seen to be associated with a large temporal change in the quantity of utterances or the power balance and behavioral pattern. Although this observation gives insights into the specification of critical moments, the observed result is inconclusive because of small number of samples. Further study across a wider range of participants and other kinds of topics is needed.

References

Dehaene, S., Spelke, E., Pinel, P., Stanescu, R., Tsivkin, S. (1999). Sources of mathematical thinking: Behavioral and brain-imaging evidence. *Science*, *284*, 970-974.

Ishikawa, N., Okazawa, T., Egi, H. (2019). Diana-AD: Dialogue analysis system for adjusting duration on real time in discussion. *Lecture Notes in Computer Science*, 11677, 212-221.

 © ISLS

Lakoff, G., & Nunez. R. E. (2001) *Where mathematics comes from: How the embodied mind brings mathematics into being*. Basic Books.

Sfard, A. (2008). *Thinking as communicating: Human development, the growth of discourses, and mathematizing*. Cambridge University Press.

Schwarz, B. B., Prusak, N., Swidan, O., Livny, A., Gal, K., Segal, A. (2018). Orchestrating the emergence of conceptual learning: a case study in a geometry class. *International Journal of Computer-Supported Collaborative Learning, 13*, 189-211.

Acknowledgments

This work was supported by JSPS KAKENHI 21K02752, 22K02951, and 25K06477.

 © ISLS

International Society of the Learning Sciences

Leveraging CSCL for Social Emotional Learning: Early Insights from a Large-Scale Classroom Study

Jiangang Hao, Patrick Kyllonen, Emily Kerzabi, Wenju Cui, Enruo Guo, Jessica Andrews-Todd, Yang Jiang
jhao@ets.org, pkllyonen@ets.org, ekerzabi@ets.org, wcui@ets.org, eguo@ets.org, jandrewstodd@ets.org, yjiang002@ets.org
ETS Research Institute

Stephen Elliott, Arizona State University, steve@ssiscolab.com

Abstract: Social and Emotional Learning (SEL) helps students to develop healthy identities, regulate emotions, build relationships, and make responsible decisions. Effective SEL instruction often relies on group-based activities; however, traditional classrooms without technological support make it difficult to capture the fine-grained interactions among students. As a result, teachers struggle to provide timely feedback and interventions. This paper reports on an ongoing project that leverages a computer-supported collaborative learning (CSCL) environment to implement 36 SEL lessons with fifth and sixth-grade students over the course of a semester. We describe the design and implementation of the curriculum and present preliminary findings from a pilot study. Early results indicate positive feedback from both teachers and students, along with a negative correlation between students' off-topic chat behavior, characterized by nonsensical or "gibberish" messages, and their learning performance.

Introduction

Social and Emotional Learning (SEL) is a fundamental component of education that helps students develop essential life skills, including emotional regulation, empathy, relationship-building, and responsible decision-making. These competencies are crucial for developing healthy identities and navigating the complexities of social interactions both in and outside of school (Bierman et al., 2008; Durlak et al., 2010; Durlak et al., 2011; Durlak et al., 2022; Greenberg, 2023; Nelson et al., 2003; Domitrovich et al., 2017; Taylor et al., 2017). Teaching SEL presents various challenges due to the skills being rooted in interpersonal engagement and collaboration. The CASEL framework (Borowski, 2019; Jagers et al., 2019) proposes that students develop five competencies: self-awareness, self-management, social awareness, relationship skills, and responsible decision-making. Assessing students' SEL competencies has traditionally relied on self-report questionnaires and observational rating scales. While these methods provide useful summaries, they offer limited insight into the dynamic processes through which students interact, communicate, and collaborate to develop SEL skills in real-world contexts (McKown, 2019).

To address this challenge, we adopted a well-established, evidence-based Social Skills Improvement System Classwide Intervention Program (SSIS CIP; Elliott & Gresham, 2007; Elliott et al., 2022) and integrated 36 lessons into a computer-supported collaborative learning (CSCL) system, the ETS Platform for Collaborative Assessment and Learning (EPCAL; Hao et al., 2017), see Figure 1 for the teacher and student interfaces. More detailed design considerations can be found in (Guo et al., in press). CSCL has been extensively shown to enhance both students' academic learning outcomes and social skills by engaging learners in joint problem-solving processes (Stahl, Koschmann, & Suthers, 2006; Jeong, Hmelo-Silver, & Jo, 2019). Within each class, students use Chromebooks to log in to the EPCAL system, where they are assigned by the teacher into small teams of two or three to collaboratively engage with the curriculum. The platform delivers the learning tasks and collects student responses and their communication process data. Using advanced analytics, the system provides teachers with actionable insights into each student's learning processes, enabling targeted interventions and continuous feedback. Moreover, team composition changes between lessons by design, providing a unique opportunity to examine how students' collaboration patterns change when working with different peers.

The full study is an efficacy study designed to evaluate the differences in learning outcomes between students participating in the computer-supported collaborative SEL curriculum and those in a control group. To fulfill this goal, we will implement the 36-lesson curriculum to 400 students over a semester, using a multi-site cluster-randomized trial with treatment assigned at the classroom level. Ten schools are participating, each contributing eight classrooms—four assigned to treatment (two 5th-grade and two 6th-grade) and four serving as controls, resulting in a total of 80 classrooms. Within each classroom, approximately 10 students with parental

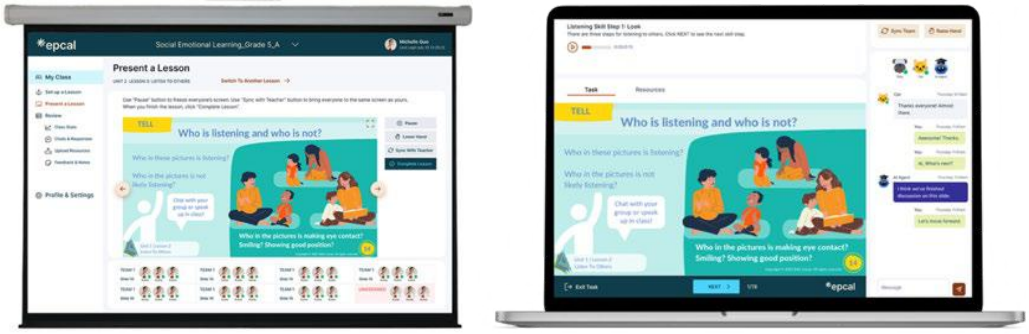
 © ISLS

consent will participate in data collection, resulting in 400 students in the treatment group and 400 in the "business-as-usual" control group.

Prior to the full study, we conducted a pilot from February to June 2025 in Michigan (U.S.) schools with 23 students and two teachers who completed 54 lessons. The pilot aimed to gather teachers' and students' perceptions of the online collaborative SEL curriculum and to identify areas for improving the system and its implementation. In this short paper, we present the overall curriculum design, summarize teacher and student feedback from the pilot, and highlight an interesting finding that gibberish chats are negatively associated with student learning—consistent with prior evidence that off-task or disruptive talk in collaborative learning environments can hinder engagement and learning outcomes (Carpenter et al., 2020; Park et al., 2021).

Figure 1

Teacher (left) and Student (right) Interfaces



In addition, both teachers and students have access to dashboards that provide summaries of the learning activities. The teacher dashboard displays students' communication quality (e.g., on-topic discussions vs. off-task chats), responses to questions, and patterns of interaction within teams. It also enables teachers to provide personalized feedback to individual students. The student dashboard presents information, including their own contributions compared to those of other team members, teacher feedback, and badges awarded by teachers.

Preliminary findings from pilot study

Reception by teachers and students

Teachers reported positive perceptions of EPCAL, highlighting ease of use, engaging content, and smooth implementation, with minor concerns about some role-play scenarios. Most students found the activities enjoyable and beneficial, especially for collaboration, while a few suggested improvements such as adding a back button, refining chat features, and simplifying certain tasks.

Gibberish message and learning performance

We analyzed 2,587 chat turns from the pilot, coding each message as gibberish (1) or not (0). We defined a Gibberish Index (GI) combining (a) the fraction of gibberish turns and (b) cross-session prevalence (proportion of sessions with any gibberish), computed as their product. Student performance was measured using two embedded questions for each lesson: The first was a self-reported rating, in which students evaluated their grasp of the targeted skills on a scale of 0 to 4 after completing each lesson. The second was a situational judgment question, which assessed whether students could select the most appropriate response in a scenario-based situation based on what they learned in that lesson, scored as 0 or 1 .For each student, we averaged scores across completed sessions and summed them into a total score. Pearson correlation showed a significant negative relationship between GI and performance ($r = -0.419$, $p = 0.0467$), suggesting more gibberish is associated with lower learning. Given the small pilot sample, results are preliminary and will be validated with full data.

Summary

Preliminary findings indicate positive responses from both teachers and students, as well as valuable insights from the process data captured by the EPCAL platform. Initial analyses of the interaction process data suggest that meaningful patterns exist linking students' communication behaviors to their learning outcomes. While these results are preliminary, they demonstrate the promise of combining digital collaborative environments with fine-grained process data to advance the assessment of engagement and SEL in authentic classroom contexts. Findings from the full study will be reported in the future.

References

Bierman, K. L., Domitrovich, C. E., Nix, R. L., Gest, S. D., Welsh, J. A., Greenberg, M. T., ... & Gill, S. (2008). Promoting academic and social-emotional school readiness: The Head Start REDI program. *Child development*, *79*(6), 1802-1817.

Borowski, T. (2019). CASEL's framework for systemic social and emotional learning. *Measuring SEL: Using Data to Inspire Practice*, *8*, 1-7.

Carpenter, D., Emerson, A., Mott, B. W., Saleh, A., Glazewski, K. D., Hmelo-Silver, C. E., & Lester, J. C. (2020, June). Detecting off-task behavior from student dialogue in game-based collaborative learning. In International conference on artificial intelligence in education (pp. 55-66). Cham: Springer International Publishing.

Domitrovich, C. E., Durlak, J. A., Staley, K. C., & Weissberg, R. P. (2017). Social-emotional competence: An essential factor for promoting positive adjustment and reducing risk in school children. *Child development*, *88*(2), 408–416.

Durlak, J. A., Weissberg, R. P., & Pachan, M. (2010). A meta-analysis of after-school programs that seek to promote personal and social skills in children and adolescents. *American journal of community psychology*, *45*(3), 294–309.

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child development*, *82*(1), 405–432.

Durlak, J. A., Mahoney, J. L., & Boyle, A. E. (2022). What we know, and what we need to find out about universal, school-based social and emotional learning programs for children and adolescents: A review of meta-analyses and directions for future research. *Psychological Bulletin*, *148*(11-12), 765.

Elliott, S. N., & Gresham, F. M. (2013). Social skills improvement system. In Encyclopedia of autism spectrum disorders (pp. 2933–2935). Springer, New York, NY.

Elliott, S.N., Malecki, C.K., Demaray, M.K., & Gresham, (2022). SSIS SEH Classwide Intervention Program: Training, Implementation, & Evaluation. Scottsdale, AZ: SAIL CoLab.

Greenberg, M. T. (2023). Evidence for Social and Emotional Learning in Schools. *Learning Policy Institute*.

Guo, E., Kerzabi, E., Andrews-Todd, J., Hao, J., & Kyllonen, P. (in press). Designing digital enhancements for social emotional skill in a collaborative learning environment. Paper accepted to the International Conference on Computer-Supported Collaborative Learning (CSCL 2026)

Hao, J., Liu, L., von Davier, A. A., Lederer, N., Zapata-Rivera, D., Jakl, P., & Bakkenson, M. (2017). EPCAL: ETS platform for collaborative assessment and learning. *ETS Research Report Series*, *2017*(1), 1–14.

Jagers, R. J., Rivas-Drake, D., & Williams, B. (2025). Transformative social and emotional learning (SEL): Toward SEL in service of educational equity and excellence. In *Social and Emotional Learning* (pp. 39–62). Routledge.

Jeong, H., Hmelo-Silver, C. E., & Jo, K. (2019). Ten years of Computer-Supported Collaborative Learning (CSCL): A meta-analysis of CSCL in STEM education during 2005–2014. Educational Research Review, 28, 100284.

McKown, C. (2019). Challenges and opportunities in the applied assessment of student social and emotional learning. *Educational Psychologist*, *54*(3), 205–221.

Nelson, G., Westhues, A., & MacLeod, J. (2003). A meta-analysis of longitudinal research on preschool prevention programs for children. *Prevention & Treatment*, *6*(1), 31a.

Park, K., Sohn, H., Mott, B., Min, W., Saleh, A., Glazewski, K., ... & Lester, J. (2021, April). Detecting disruptive talk in student chat-based discussion within collaborative game-based learning environments. In LAK21: 11th International Learning Analytics and Knowledge Conference (pp. 405-415).

Stahl, G., Koschmann, T., & Suthers, D. (2006). Computer-supported collaborative learning: An historical perspective. In R. K. Sawyer (Ed.), *Cambridge handbook of the learning sciences* (pp. 409–426). Cambridge University Press.

Taylor, R. D., Oberle, E., Durlak, J. A., & Weissberg, R. P. (2017). Promoting positive youth development through school-based social and emotional learning interventions: A meta-analysis of follow-up effects. *Child development*, *88*(4), 1156–1171.

Acknowledgments

This work was funded in part by the Education Innovation Research (EIR) program of the Office of Elementary and Secondary Education (OESE), U.S. Department of Education, (#S411C230179). The views, opinions, and/or findings contained in this paper are those of the authors and shall not be construed as an official position, policy, or decision of the funding agencies, unless so designated by other documents.

International Society of the Learning Sciences

Collaborative High School Teacher Learning in Bichronous Online STEM Professional Development

Ho Sue Kwon, University of Pennsylvania, suekwon@upenn.edu
Susan A. Yoon, University of Pennsylvania, yoonsa@upenn.edu
Jooeun Shim, Teachers College, Columbia University, js4719@tc.columbia.edu

Abstract: This study examines how a bichronous model supports collaborative teacher learning in a four-week online bioinformatics professional development. Through analysis of nine high school teachers' 1,363 asynchronous forum posts and 220 minutes of synchronous sessions, we identified complementary pathways: asynchronous forums supported knowledge building through archived threads and peer resource sharing, while synchronous sessions facilitated real-time collective sensemaking and expert scaffolding. Findings offer design principles for online teacher professional development in emerging STEM domains.

Introduction

Designing online teacher professional development (oTPD) to support both sustained reflection and real-time collaboration within time constraints is a fundamental challenge (Haug & Mork, 2021; Meyer et al., 2023; Mulaimović et al., 2024), particularly in emerging interdisciplinary STEM fields like bioinformatics (Yoon et al., 2022). Without standardized curricula, teachers need to acquire both bioinformatics content knowledge and pedagogical content knowledge which involve designing curricula using authentic data sets, simulations, and computational tools to engage students in scientific knowledge building in multiple domains (Huang et al., 2022; Yoon et al., 2025). oTPD models that blend delayed reflection and real-time interaction with colleagues and experts within a sustained professional learning community (PLC) help optimize teachers' use of time while meeting core principles for teacher learning for STEM integration (Jeong et al., 2019; Yoon et al., 2025).

Existing online learning environments have demonstrated the potential of virtual space to transform teacher learning, yet the field lacks instructional models showing how to orchestrate multiple online modalities for meaningful oTPD experiences (Bragg et al., 2021; Jin et al., 2024). We conceptualize bichronous learning as the purposeful integration of asynchronous and synchronous modalities to sustain engagement in fully oTPD (Martin et al., 2023). Asynchronous spaces offer temporal flexibility, access to an accumulating archive, and opportunities for iterative reflection and peer-curated resources (Clark & Mayer, 2024; Hernandez et al., 2024; Moore et al., 2024). Synchronous sessions afford immediacy for real-time collective problem-solving, just-in-time expert feedback and modeling, and community building that generates momentum and accountability (Clark & Mayer, 2024; Jin et al., 2024; Mulaimović et al., 2024). We ask, how does a bichronous model support collaborative learning as teachers develop expertise in emerging STEM domains?

Methodology

The bichronous model was instantiated on edX Edge in July 2020. The asynchronous layer used threaded forums with search functions and embedded multimedia. It allowed teachers to initiate conversations, revisit prior contributions, and build shared resource collections. The synchronous layer employed integrated video sessions with screen sharing for collaborative problem-solving, facilitator-guided Q&A with subject-matter experts, and a live chat to parallel discussion. The design blended asynchronous modules with weekly live sessions totaling approximately 60 hours of learning.

Nine carefully selected high school (9th-12th grade) biology and environmental science teachers from Northeastern US participated (IES & NSF, 2013). Their teaching experience ranged from 2 to 14 years (m = 5 years). All teacher participants completed asynchronous assignments and joined weekly synchronous meetups alongside the research team, subject matter experts, and two teacher facilitators from the previous cohort.

Two primary data sources included: (1) asynchronous discussion forum prompts (n = 52) with participant responses (n = 1,363), categorized as top-level posts, second-level comments, and third-level replies; and (2) synchronous session recordings (n = 5), including one orientation and four weekly Friday meetings. The synchronous sessions, totaling 220 minutes, included content discussions, announcements, check-ins, Q&A, and feedback.

We conducted a hybrid deductive-inductive thematic analysis (Fereday & Muir-Cochrane, 2006) of asynchronous and synchronous data that integrated theory-driven constructs from teacher PD (Darling-Hammond et al., 2017) with inductive codes grounded in teachers' interactions and practices.

 © ISLS

Findings

Analysis revealed an iterative sequence where asynchronous exploration prepared ground for synchronous synthesis, which sparked deeper problem-solving across modalities. This cyclical progression supported teachers in co-constructing bioinformatics content knowledge and pedagogical content knowledge within the virtual PLC.

Phase 1 was asynchronous foundation building. Teachers used asynchronous forums for sustained, situated reflection for developing competence in teaching bioinformatics. The temporal flexibility allowed them to explore new concepts, contextualize challenges, and connect content to students' lived experiences. For example, one teacher reflected on the local prevalence of asthma among their students, "I would estimate over 80% of my students had been diagnosed with asthma … looking into it deeper, I found this article [saying] 1 in 4 kids in Philadelphia are diagnosed with asthma [due primarily to poor air quality]." With sufficient time given, they were able to construct content knowledge by researching and drawing connections between real-world data and classroom instruction. The forum became a dynamic archive for exploring unfamiliar STEM content and sharing teaching resources contextualized for their classroom use.

Phase 2 was synchronous synthesis through expert scaffolding and peer collaboration. Facilitators strategically monitored asynchronous threads to identify emergent themes that required synchronous attention. Each synchronous session began with reflections on recent forum activity to strengthen social ties and shared purposes within the PLC. When teachers expressed instructional challenges asynchronously, facilitators addressed these in real-time with conceptual scaffolding. For instance, one teacher had expressed concerns about using problem-based learning (PBL) with students who lack a sufficient foundation of academic skills. In the synchronous discussion, a facilitator responded, "[PBL] is highly scaffolded … You're scaffolding the groups, the questions … you're embedding this in a problem that they have a choice in solving." This immediate reframing helped teachers perceive PBL as structured and adaptable and created a sense of co-presence and urgency that propelled the group toward actionable planning and implementation.

Phase 3 was collaborative problem-solving across modalities. Following synchronous clarification, teachers returned to asynchronous forums with enhanced understanding. Topics raised tentatively in Phase 1 were revisited with greater clarity and confidence. The remote learning solutions that were generated synchronously re-emerged in later asynchronous modules as teachers shared refined implementation plans, now incorporating specific strategies for data collection and student inquiry into PBL instruction. The asynchronous modality continued to afford teachers the time and autonomy to explore, synthesize, and enact new noncanonical STEM knowledge with growing competence. This ongoing cycle of inquiry, peer feedback, and facilitation, now informed by synchronous insights, established a pattern where each modality's affordances reinforced one another.

In sum, the bichronous design enabled continuous cycling between modalities. Asynchronous learning supported conceptual exploration, resource sharing, and idea generation while synchronous learning catalyzed clarification, problem-solving, and implementation. This strategic sequencing transformed tentative exploration into confident implementation planning, establishing the core pillars of high-quality PD: iterative sense-making, authentic problem-solving, and continuous PLC building.

Discussion and conclusion

Deliberate coordination across time and modality fostered a cumulative learning ecology where sustained reflection and real-time collaboration reinforce one another (Huang et al., 2022; Meyer et al., 2023; Mulaimović et al., 2024; Yoon et al., 2025). Three design principles emerge from this model. First, integrate platforms for cognitive continuity (Bragg et al., 2021; Jeong et al., 2019). Hosting discussion forums, live sessions, resources, and artifacts within one learning environment reduces context switching, preserves interactional history, and turns movement between "any-time" and "real-time" into part of the learning flow rather than a friction point. Second, leverage continuity as a collaborative scaffold (Clark & Mayer, 2024; Hernandez et al., 2024; Moore et al., 2024). A searchable, preserved archive enabled teachers to resurface ideas across weeks, to justify design choices with prior discussion, and to onboard peers into ongoing threads. Continuity supported both individual sense-making and collective progress (Haug & Mork, 2021). Third, design for complementary, not redundant affordances (Jin et al., 2024; Martin et al., 2023). Cross-modal scaffolds such as prompts that seeded live agendas or synchronous inquiries that returned to the forum transformed separate channels into one coordinated learning ecology.

Our study advances CSCL theory by shifting focus from comparing different modalities to specifying strategic coordination mechanisms. Future research should examine how the bichronous model scales in larger, more diverse teacher PLCs, how they impact on classroom enactment and student learning, and how AI might enhance these feedback loops. Platform decisions are pedagogical decisions: they shape when and where learning happens and sustain collaboration across time.

 © ISLS

References

Bragg, L. A., Walsh, C., & Heyeres, M. (2021). Successful design and delivery of online professional development for teachers: A systematic review of the literature. *Computers & Education, 166*, 104158.

Clark, R. C. & Mayer, R. E. (2024). e-Learning: Promises and pitfalls. In Clark, R. C. & Mayer, R. E., *e-Learning and the science of instruction: Proven guidelines for consumers and designers of multimedia design (5th ed.)*. Wiley.

Darling-Hammond, L., Hyler, M., & Gardner, M. (2017). *Effective Teacher Professional Development*. Learning Policy Institute.

Fereday, J., & Muir-Cochrane, E. (2006). Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *International journal of qualitative methods, 5*(1), 80-92.

Haug, B. S., & Mork, S. M. (2021). Taking 21st century skills from vision to classroom: What teachers highlight as supportive professional development in the light of new demands from educational reforms. *Teaching and Teacher Education, 100*, 103286.

Hernandez, L., Archambault, L., & Harris, L. M. (2024). Examining blended professional development in the aftermath of the COVID-19 pandemic: An action research study. *Journal of Online Learning Research 10*(2), 255-280.

Huang, B., Siu-Yung Jong, M., Tu, Y.-F., Hwang, G.-J., Chai, C. S., & Yi-Chao Jiang, M. (2022). Trends and exemplary practices of STEM teacher professional development programs in K-12 contexts: A systematic review of empirical studies. *Computers & Education, 189*, 104577.

Institute of Education Sciences (IES) & National Science Foundation (NSF). (2013). *Common Guidelines for Education Research and Development.*

Jeong, H., Hmelo-Silver, C. E., & Jo, K. (2019). Ten years of Computer-Supported Collaborative Learning: A meta-analysis of CSCL in STEM education during 2005–2014. *Educational Research Review, 28*, 100284.

Jin, F., Song, Z., Cheung, W. M., Lin, C. H., & Liu, T. (2024). Technological affordances in teachers' online professional learning communities: A systematic review. *Journal of Computer Assisted Learning, 40*(3), 1019-1039.

Martin, F., Kumar, S., Ritzhaupt, A. D., & Polly, D. (2023). Bichronous online learning: Award-winning online instructor practices of blending asynchronous and synchronous online modalities. *The Internet and Higher Education, 56*, 100879.

Meyer, A., Kleinknecht, M., & Richter, D. (2023). What makes online professional development effective? The effect of quality characteristics on teachers' satisfaction and changes in their professional practices. *Computers & Education, 200*, 104805.

Moore, B. A., Legleiter, E. F., Owens, K., Packard, B., & Wright, J. (2024). Evaluating the Effectiveness of an Online Professional Development for Rural Middle-School Science Teachers. *Journal of Science Teacher Education*, *35*(4), 386–404.

Mulaimović, N., Richter, E., Lazarides, R., & Richter, D. (2024). Comparing quality and engagement in face-to-face and online teacher professional development. *British Journal of Educational Technology*.

Yoon, S. A., Miller, K. M., Richman, T., Noushad, N., Hageman, G., Liu, Y., Zhang, W., & Cottone, A. M. (2025). Developing teachers' disciplinary knowledge for high school STEM integration: A review of a decade of educational research. *Review of Educational Research*, *95*(6), 1337–1378.

Yoon, S. A., Shim, J., Miller, K., Cottone, A. M., Noushad, N. F., Yoo, J., Gonzalez, M. V., Urbanowicz, R., & Himes, B. E. (2022). Professional development for STEM integration: Analyzing bioinformatics teaching through considerations of adaptive expertise and cognitive load. In Superfine, A.C., Goldman, S. R., & Ko, M. M. (Eds.), *Teacher learning in changing contexts: Perspectives from the learning sciences* (pp. 69–90). Routledge.

Acknowledgments

This work was supported by a grant from the U.S. National Science Foundation (DRL #1812738). We thank Siyun Chen and Rainy Phrompechrut for their assistance in developing codebooks for data analysis.

International Society of the Learning Sciences

Enhancing K-12 Teachers' Confidence and Competence in STEM Integration Using Blended Learning in Professional Development

Haowei Li, Amanda M. Cottone, Susan A. Yoon, Zheng Bian, Rebecca G. Wells
lhaowei23@gmail.com, amandaco@upenn.edu, yoonsa@upenn.edu, maxbian0925@gmail.com, rgwells@pennmedicine.upenn.edu
University of Pennsylvania

Abstract: STEM-integrated learning in K-12 education is essential for students to build career readiness. However, building confidence and competence in STEM teachers to enact new, cross-disciplinary curriculum is difficult, and the best way to design high-quality professional development (PD) programs for teaching STEM integration is unclear. This study introduces a blended learning PD model that added a new online course to a traditional onsite research experience to scaffold teachers' comprehension of *mechanobiology* knowledge and practices of pedagogy. Piloting the model with seven K-12 teachers, we examined how this blended format scaffolded in-person and online collaborative learning. Findings show the online course enhanced teachers' grasp of complex research concepts and competence in teaching STEM-integrated topics. This study highlights the potential of an online course to promote confidence and competence in teachers conducting cutting-edge, interdisciplinary STEM research.

Introduction

Fostering student engagement in STEM via real-world research contexts is necessary to strengthen workforce preparation (NRC, 2012). This requires K-12 educators to shift towards STEM integration, applying cross-disciplinary knowledge to solve complex problems (Schweingruber et al., 2014). It further requires an emphasis on engineering design practices to enhance students' ability to apply interdisciplinary knowledge (NGSS Lead States, 2013; Moore et al., 2014). Despite the interdisciplinary nature of modern STEM, K-12 education has been slow to keep pace (Yoon et al. 2024; Caner & Ogan-Bekiroglu, 2025). Targeted PD for STEM-integrated topics must support gains in two essential constructs: *confidence*, or self-efficacy in delivering curricula, and *competence*, the Content Knowledge (CK) and Pedagogical Content Knowledge (PCK) required to enact it (Acar & Büyüksahin, 2021; Caner & Ogan-Bekiroglu, 2025). Research Experiences for Teachers (RETs) are effective mechanisms for this, offering in-person, collaborative environments with STEM researchers (Yoon et al., 2024). However, the optimal design features of RETs vary widely and are not fully understood (Enderle et al., 2014). To address these challenges, we propose a blended learning model, leveraging technologies and online platforms to add flexibility and tailored content, amplifying learning outcomes (Tabak & Kyza, 2018; Tayebinik & Puteh, 2013). Our theoretical mechanism posits that asynchronous online learning builds foundational CK by utilizing multimedia assets to facilitate understanding via multiple cognitive channels. This allows in-person time to be dedicated to high-quality PCK development and collaboration. We focused this RET on mechanobiology, a field merging physics, engineering, and biology, providing an authentic context for interdisciplinary STEM integration.

Methods

We piloted the blended design in a six-week mechanobiology RET in 2025 at a northeastern U.S. research center. Participants were seven local K-12 teachers representing diverse STEM subjects and teaching experience levels (ranging from 3 to 17 years). Teachers worked with a graduate lab mentor on a research project while completing three online course modules. The online modules included multimedia assets, peer discussion boards to establish peer support, discipline-specific scaffolds, examples of real-world mechanobiology applications, and expert lectures to be paired with in-person flipped classrooms. We collected pre- and post-program surveys, online coursework (lesson plans, discussion entries), and transcripts from three focus groups. We analyzed open-ended surveys and transcripts using interaction analysis to investigate collaboration. We employed a modified grounded theory approach to evaluate 114 minutes of focus group audio, iteratively searching for evidence of teachers connecting their learning to specific design features (Glaser, 2008).

Results

The results section is organized around two key findings: 1) the blended PD model supported teachers' build of confidence and competence in STEM-integrated teaching, and 2) specific design features, including examples of real-world applications and opportunities for collaboration, were instrumental in this change.

 © ISLS

Teachers report grasp of CK after computer-supported learning in PD

Six of the seven teachers completed the post-program survey and indicated high levels of confidence (on a 1–5 scale, with 1 being the lowest) in their understanding of "what mechanobiology is", "why it matters" (Mean = 4.67), and "how physical forces, such as stress and strain, affect cells and tissues" (Mean = 4.33). These results reflect consistently strong confidence in CK related to mechanobiology, which coincides with the evidence of a teacher explicitly stating in a focus group session that he gained confidence after the blended RET, mentioning "I am the expert in mechanobiology in the room now." Five out of the six teachers agreed that they gained confidence in "implementing hands-on or project-based learning experiences in mechanobiology with [their] students" (Mean = 4.2) and in "teaching interdisciplinary STEM to [their] students" (Mean = 4.2), indicating an increase in confidence of PCK in STEM integration. Four teachers indicated in the post-survey that they understood the core principles of mechanobiology and could connect them to their teaching after completing the online course.

The 2025 cohort stated that the online course included foundational concepts in mechanobiology and STEM integrated knowledge that led to elevated confidence in completing lab work. Speaking of his experience participating in the RET for the second year, a teacher compared the in-person program in 2024 with the new blended design added in 2025: "[The online course] really helped me apply [mechanobiology content knowledge] in the lab this year, because last year I was a little lost, and this year, I have a much better grasp on things." Completing the online course concurrently with lab research also benefited first year teachers in building a foundation of knowledge in mechanobiology. One first time participant said that learning about mechanobiology from a blended learning environment was helpful: "I think [the online course] helped just give…background information and a foundation…as the summer went on…I felt like I was understanding things better and better." Another first year RET participant reported that online course content prepared her to participate in lab meetings: "we're in these…intimidating lab meetings, with everybody and the PIs…I liked the fact that I knew what [mechanosensing proteins] were prior to entering these meetings…it was good just having that constant repetition from this course." A teacher further explained "It's pretty overwhelming when you start in the lab with all these, you know, new machines, new people…So it was really helpful to have a solid grasp on science concepts first." This suggests new and returning teachers agreed that the online course strengthened their understanding of complex, mechanobiology knowledge and promoted confidence in lab research.

Design features of the online course towards developing teachers' competence

Key design features for developing competence in STEM-integration included real-life applications, classroom examples, and expert access, which were implemented with multimedia assets including lab and classroom videos, podcasts, images, text, hands-on lesson plans, student worksheets, and teacher guides. Teachers showed gains in CK, shifting from general to applied understandings of mechanobiology. For instance, one teacher's definition of mechanobiology evolved from an "interdisciplinary of [sic] biological systems" (pre-survey) to "understanding diseases, helping cure and prevent such diseases" (post-survey). Four of six teachers ranked interacting with "actual research and innovations" as the most useful feature, alongside multimedia assets, peer discussion boards, discipline-specific scaffolds, and real-world application examples. The application-based approach also supported PCK development with teachers preparing lesson plans using specific industry skincare examples to inspire classroom activities and captivate students uninterested in the traditional curriculum. Additionally, combining online expert resources with in-person "flipped classroom" meetups established a social accountability structure that deepened content engagement and collaborative PCK reflection more effectively than self-directed modules alone. For instance, teachers noted that viewing an expert's classroom enactment and subsequent discussions made them reflect on their instruction and explained the accountability stimulated engagement: "when I would see these hour-long lectures, I was bought in, because I'm meeting these people tomorrow." Finally, while the experience of being a student boosted their confidence, teachers exhibited varying preferences for how CK and PCK should be sequenced. Some prefer to build deep conceptual understanding before practicing pedagogy, and others favor concurrent learning, indicating a need for flexible online professional development structures.

Discussion

This pilot study advances the field by providing a scalable blueprint for research-based PD. Our findings demonstrate that utilizing online modules to manage the complexity of STEM-integrated CK allows teachers to reserve in-person interactions for deep PCK reflection and collaboration (Acar & Büyüksahin, 2021; Caner & Ogan-Bekiroglu, 2025). This addresses the core requirement of the NGSS – a focus of engineering design process – and responds to the need to bring engineering practices to biology classes (Lead States, 2013). Future research will focus on modularizing dense content to offer more flexible pathways between deep content acquisition and integrated pedagogical reflection.

References

Acar, D., & Büyüksahin, Y. (2021). Awareness and Views of Teachers Who Received In-Service STEM Training about STEM. *International Journal of Progressive Education*, *17*(2), 473-490.

Caner, F., & Ogan-Bekiroglu, F. (2025). Determination of Integrated STEM Teacher Competencies: A Modified Delphi Method. *European Journal of STEM Education*, *10*(1), 06. https://doi.org/10.29333/ejsteme/14134

Enderle, P., Dentzau, M., Roseler, K., Southerland, S., Granger, E., Hughes, R., ... & Saka, Y. (2014). Examining the influence of RETs on science teacher beliefs and practice. *Science Education*, *98*(6), 1077-1108. https://doi.org/10.1002/sce.21127

Moore, T. J., Glancy, A. W., Tank, K. M., Kersten, J. A., Smith, K. A., & Stohlmann, M. S. (2014). A framework for quality K-12 engineering education: Research and development. *Journal of Pre-College Engineering Education Research*, *4*(1), 02. https://doi.org/10.7771/2157-9288.1069

National Research Council (NRC). (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. National Academies Press. https://doi.org/10.17226/13165

NGSS Lead States. (2013). *Next generation science standards: For states, by states*.

Schweingruber, H., Pearson, G., & Honey, M. (Eds.). (2014). *STEM integration in K-12 education: Status, prospects, and an agenda for research*. National Academies Press. https://doi.org/10.17226/18612

Tabak, I., & Kyza, E. A. (2018). Research on scaffolding in the learning sciences: A methodological perspective. In *International handbook of the learning sciences* (pp. 191-200). Routledge.

Tayebinik, M., & Puteh, M. (2013). Blended Learning or E-learning?. *arXiv preprint arXiv:1306.4085*. https://doi.org/10.48550/arXiv.1306.4085

Yoon, S. A., Miller, K. M., Richman, T., Noushad, N., Hageman, G., Liu, Y., ... & Cottone, A. M. (2024). Developing Teachers' Disciplinary Knowledge for High School STEM Integration: A Review of a Decade of Educational Research. *Review of Educational Research*, 00346543241289566.

Acknowledgments

This work was supported by the Center for Engineering MechanoBiology (CEMB), an NSF Science and Technology Center, under grant agreement CMMI: 15-48571.

International Society of the Learning Sciences

Knowledge-Building Discourse in Generative AI-Mediated Collaborative Problem Solving: A Student–Expert Comparison

Bunichi Otaki, Ritsuko Oshima, Jun Oshima
otaki.bunichi.23@shizuoka.ac.jp, roshima@inf.shizuoka.ac.jp, joshima@inf.shizuoka.ac.jp
Shizuoka University, Graduate School of Science and Technology

Abstract: This study examines how students and experts differ in their engagement with generative AI (GenAI) during knowledge-creating dialogue. Twenty undergraduate students (10 pairs) and one pair of expert educators collaborated with ChatGPT-4o to address a socio-scientific issue. Using the Knowledge Building Discourse Explorer (KBDeX) and "Good Moves" coding, we analyzed temporal degree centrality and discourse networks to capture epistemic and dialogic processes. Results showed that students' discourse was largely anchored to AI-generated ideas, with limited transformation through dialogue, whereas experts engaged in self-sustained cycles of problem definition, evaluation, and reframing, using GenAI selectively to extend inquiry. These findings suggest that differences lie not in the availability of ideas but in how discourse is organized to support knowledge creation, highlighting the importance of scaffolding epistemic and dialogic practices for effective GenAI use.

Introduction

Collaborative problem solving (CPS) in the knowledge age involves advancing ideas through sustained dialogue. Knowledge Building (KB) frames CPS as collective inquiry in which participants iteratively define and refine ideas through discourse (Bereiter & Scardamalia, 2014, 2016). Generative AI (GenAI) can support this process by providing ideas, summaries, and explanations (Essel et al., 2024), yet its effectiveness depends on whether learners critically evaluate and integrate its outputs into ongoing inquiry (Du, 2025), consistent with research emphasizing reflective dialogue and shared evaluation in CPS (Zheng et al., 2025). Students and educators differ in how they perceive GenAI's role (Otaki et al., 2026), suggesting that such differences may shape their engagement in collaborative settings. Prior research has mainly examined the outcomes or structures of AI-supported argumentation, leaving less understood how knowledge-building discourse unfolds as a process in interaction with GenAI.

From a KB perspective, expertise is understood as progressive problem solving rather than accumulated knowledge (Scardamalia & Bereiter, 1994). Comparing students and expert educators provides a lens to examine how different epistemic frames shape engagement with GenAI. Accordingly, this study addresses the following research question: How do temporal and dialogical structures of discourse differ between student groups and expert educators when engaging in knowledge-creating dialogue with GenAI?

Methods

Participants included 20 undergraduate students (10 pairs) and two expert educators (one pair). All participants collaborated with ChatGPT-4o on a shared monitor to address a socio-scientific issue related to genetically modified foods (GMOs). Students reported prior experience using ChatGPT for academic support, while the experts, who specialized in science education, reported extensive experience with inquiry-based teaching and daily use of multiple GenAI tools. Each pair engaged in a 40-minute discussion followed by a brief presentation and a short reflective interview. Data sources included audio–video recordings, screen captures of AI interactions, and transcripts of human and AI utterances segmented by conversational turn.

To examine epistemic and dialogic engagement with GenAI, we analyzed discourse using the Knowledge Building Discourse Explorer (KBDeX; Oshima et al., 2012). Words from participants' discourse were used as nodes, and their temporal total degree centrality (TDC) was calculated across turns to capture how ideas gained prominence over time. TDC was interpreted as an indicator of epistemic focus within evolving discourse. Discourse was coded for "Good Moves" (Bereiter & Scardamalia, 2016), including problem definition, new ideas, promisingness evaluation, critical discourse, comparison, meta-dialogue, and higher-level ideas. Coding was conducted at the level of conversational turns, and distributions of moves were compared between student and expert groups to examine differences in how discourse was organized and how GenAI was incorporated into knowledge-building processes.

 © ISLS

Findings and discussion

The analysis revealed systematic contrasts in how students and experts engaged in argumentation and knowledge-creating dialogue with GenAI. Figure 1 compares temporal degree centrality and discourse networks between student and expert groups.

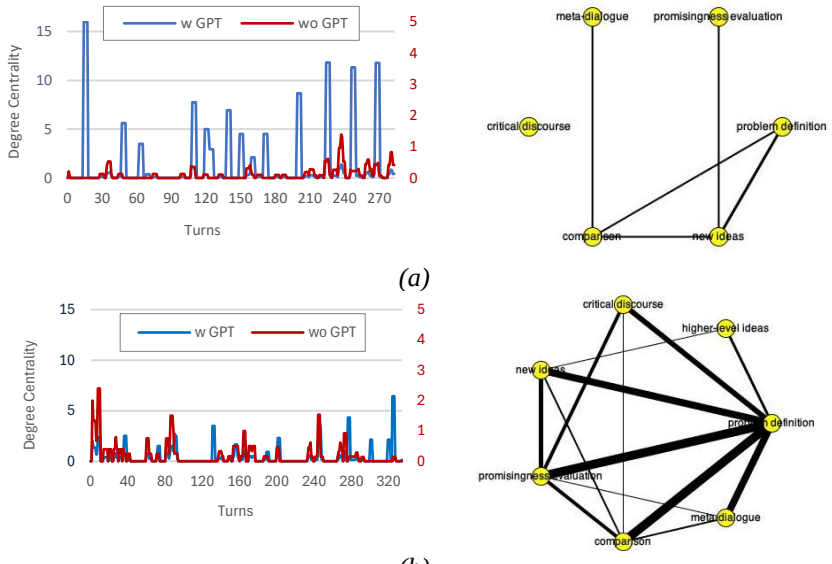
First, temporal trajectories of degree centrality (DC) showed differences in how ideas were introduced and developed. In the student group, DC values for AI-generated words increased sharply following prompts, indicating that GenAI outputs quickly became central to the discussion. Many of these ideas were incorporated directly into the evolving discourse, suggesting that students' engagement was largely anchored to AI-generated content rather than being progressively transformed through dialogue. This pattern reflects a mode of interaction in which GenAI functions as a primary driver of idea development. In contrast, the expert group showed a different pattern: increases in DC were more strongly associated with ideas introduced through their own dialogue, with GenAI used selectively to extend, refine, or consolidate emerging ideas, particularly in later phases of the discussion. Here, GenAI functioned as a secondary resource embedded within an ongoing process of inquiry rather than as a source of direction.

Second, knowledge-building discourse networks revealed qualitative differences in how dialogue was organized. The student network was relatively sparse and centered around a limited set of connections, with few links to critical discourse or meta-dialogue. This suggests that while students incorporated AI-generated ideas, these were not extensively reworked through collective inquiry, often remaining at the level of elaboration rather than transformation. In contrast, the expert network was dense and more evenly distributed, characterized by frequent connections among problem definition, critical discourse, and meta-dialogue. This pattern indicates sustained cycles of inquiry in which ideas were continuously examined, reframed, and evaluated, reflecting a more deliberate and strategically coordinated discourse.

Taken together, these findings suggest that the key difference between students and experts lies not in the availability or quality of ideas but in how discourse is organized to support knowledge creation. Students tended to engage with GenAI as a source of ideas, resulting in externally anchored discourse, whereas experts used GenAI as a resource within internally regulated inquiry, embedding AI outputs into iterative cycles of problem framing and evaluation. This distinction highlights that effective use of GenAI in CPS depends on epistemic and dialogic practices that enable learners to critically integrate AI-generated content into sustained knowledge-building processes, rather than relying on AI outputs as ready-made solutions. Pedagogically, this implies the need to design learning environments that scaffold not only access to AI-generated ideas but also the dialogic practices required to interrogate, connect, and refine those ideas over time.

Figure 1

Comparative Patterns of Discourse and Knowledge-Creating Dialogue Patterns. Student Group (a) Expert Group (b)



Note. Temporal degree centrality trajectories (left) and "Good Moves" discourse networks (right) for student and expert groups.

 © ISLS

References

Bereiter, C., & Scardamalia, M. (2014). Knowledge building and knowledge creation: One concept, two hills to climb. In *Knowledge creation in education* (pp. 35-52). Springer.

Bereiter, C., & Scardamalia, M. (2016). "Good Moves" in knowledge-creating dialogue. *QWERTY-Interdisciplinary Journal of Technology, Culture and Education*, *11*(2), 12-26.

Du, X., Du, M., Zhou, Z., & Bai, Y. (2025). Facilitator or hindrance? The impact of AI on university students' higher-order thinking skills in complex problem solving. *International Journal of Educational Technology in Higher Education*, *22*(1), 39.

Essel, H. B., Vlachopoulos, D., Essuman, A. B., & Amankwa, J. O. (2024). ChatGPT effects on cognitive skills of undergraduate students: Receiving instant responses from AI-based conversational large language models (LLMs). *Computers and Education: Artificial Intelligence*, *6*, 100198.

Oshima, J., Oshima, R., & Matsuzawa, Y. (2012). Knowledge Building Discourse Explorer: a social network analysis application for knowledge building discourse. *Educational Technology Research and Development*, *60*(5), 903-921.

Otaki, B., Naganuma, S., Jin, T., & Oshima, J. (2026). Differences in perceptions of generative AI feedback by non-native English-speaking students and educators in higher education. *Educational Psychology*, *46*(1), 112-149.

Scardamalia, M., & Bereiter, C. (1994). Computer support for knowledge-building communities. *The Journal of the Learning Sciences*, *3*(3), 265-283.

Zheng, L., Shi, Z., Gao, L., & Fan, Y. (2025). A novel group cognitive graph approach for improving collaborative problem solving. *International Journal of Educational Technology in Higher Education*, *22*(1), 18.

 © ISLS

International Society of the Learning Sciences

Integrating AI and Interaction Analysis: A Mixed-Methods Approach to Tracing Object-Based Participation in CSCL

Hakeoung Hannah Lee, University of Virginia, hannahlee@virginia.edu
Miguel Lujan, United States Department of War, mike@miguellujan.com
Venkatesh Jatla, Fralin Biomedical Research Institute at VTC, vjatla@vtc.vt.edu
Ugesh Egala, Independent Researcher, ugeshe@gmail.com
Marios Pattichis, University of New Mexico, pattichi@unm.edu
Sylvia Celedon-Pattichis, University of Texas at Austin, sylvia.celedon@austin.utexas.edu

Abstract: This study introduces and employs a novel mixed-methods design that integrates computer vision, optical character recognition (OCR), and qualitative video-based interaction analysis to examine multimodal participation in Computer-Supported Collaborative Learning (CSCL), with a focus on typing as a key form of object-based interaction. This approach integrates two computational tools: a neural network–based system for mapping typing activity in group interaction videos and a screen recording analysis system for extracting typed input. Analyzing a collaborative robotics programming session among middle school students, the findings identify who typed, when, and what was typed, as well as dynamic shifts in participation, tool access, and role negotiation. The study contributes to more inclusive, scalable and grounded analyses of CSCL processes by bridging Artificial Intelligence-enhanced automation with human insight, surfacing patterns that warrant deeper qualitative investigation.

Introduction

Understanding CSCL processes requires attention to the complex, multimodal nature of student participation (Martinez-Maldonado et al., 2021; Ouyang et al., 2023; Worsley et al., 2021). Multimodal participation dynamics, as conceptualized by Lee et al. (2026), refer to the dynamic interplay of these diverse modes of contribution as students engage with peers, tools, and artifacts within learning environments. Recognizing nonverbal and object-based forms of participation is critical for valuing the contributions of diverse students (Lee et al., 2026; Miller et al., 2022; Sinclair & de Freitas, 2019; Yeh et al., 2025). Video-based interaction analysis (Erickson, 2012; Jordan & Henderson, 1995) offers rich insights into multimodal participation dynamics but is highly labor-intensive. Recent advances in multimodal learning analytics (MMLA) offer promising tools for capturing and analyzing diverse data streams. However, many existing approaches remain costly, intrusive, and difficult to deploy in real-world settings, with limited application in naturalistic environments (Cukurova et al., 2020; Worsley et al., 2021; Yan et al., 2022). Recent work (Lee, 2025; Lee et al., 2026) contributes to addressing these challenges by using real-world collaborative group interaction videos as a primary data source for MMLA. By leveraging Artificial Intelligence (AI)-based image/video analysis technologies—particularly computer vision and neural networks—these studies enabled scalable and less intrusive extraction and quantification of multimodal human activities in authentic CSCL environments. For instance, Lee (2025) developed an AI-powered Activity Mapping Video Analysis system designed to capture and analyze three key activities: talking, typing, and writing. Grounded in sociocultural learning theories (Engeström, 1987; Vygotsky, 1978), the current study extends this line of work by emphasizing object-based participation as a critical component of CSCL and introducing novel analytical tools and approaches. Particularly, we focus on *typing* activity associated with a shared computer within collaborative student groups to explore the multimodal, specifically object-based, participation of middle school students. Building on our Activity Mapping Video Analysis system, which uses neural network models for typing activity detection, this study introduces a novel complementary tool: a Screen Recording Video Analysis system that leverages Optical Character Recognition (OCR) to extract the content being typed on screens. Together, they enable the detection of who is typing, when, and what is being typed from large-scale video data. By further integrating these systems with qualitative interaction analysis, this paper demonstrates a scalable yet context-sensitive approach for studying multimodal participation in real-world CSCL environments. Ultimately, this study contributes to more inclusive, scalable, and grounded analyses of CSCL by bridging automation with human insight, particularly through surfacing patterns and moments that warrant deeper qualitative investigation.

Methods

The neural network models for the Activity Mapping Video Analysis system were developed using 2,218 hours of multimedia data capturing collaborative student interactions in a three-year integrated mathematics and computer science program for underrepresented middle school students in the Southwest U.S. See Lee et al. (2023,

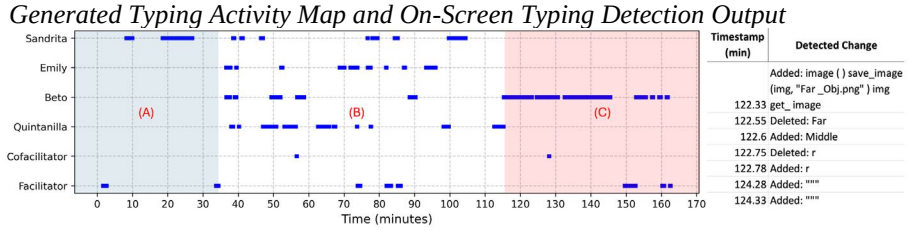
2024) for additional program details. We focus on one cohort that participated in a 12-session, three-hour daily program (June 11–29, 2018; 36 students in six groups, >85% Latine). Each group included 2–5 students, a facilitator (a STEM major (under)graduate student), and a co-facilitator (a returning middle school student). Our analysis centers on one group during a three-hour session (June 28), in which students programmed a Raspberry Pi–based robot to detect and track colored objects while exploring RGB-based inequalities and histograms for a final presentation. In Phase 1 (*Typing activity detection and visualization*), typing activity was detected and visualized in collaborative group video recordings using the AI-based Activity Mapping Video Analysis System (Lee, 2025). This followed a five-stage pipeline: (1) participant detection (Jatla et al., 2025), (2) typing classification via a 3D-CNN, (3) optimized batch inference, (4) the Human-AI Activity Mapping Cross-Validation (Lee, 2025), and (5) generation of Typing Activity Maps that visualize the temporal distribution of typing activity. The model achieved 69.58% accuracy with improved efficiency over prior methods. In Phase 2 (*Screen recording analysis*), screen recordings were analyzed using adaptive sampling and AWS Textract OCR to identify typed content. This approach achieved 87% accuracy with reduced computational load. In Phase 3 (*Qualitative examination of collaborative dynamics*), an 11.5-minute segment (1:55:00[115min]–2:06:30[126.5min]) was selected to examine shifts in participation, where one student became the sole typist (see Figure 1, Phase B→C). Video-based interaction analysis (Erickson, 2012; Jordan & Henderson, 1995) was used to analyze how students' roles evolved, how their interactions shaped the CSCL process, and to synthesize results from prior phases.

Results and discussion

The analysis revealed distinct shifts in multimodal participation throughout the collaborative programming session, particularly in how typing activity—and, by extension, access to the shared programming interface—was distributed among students. Three interaction phases emerged (see Figure 1; details visible upon magnification). The horizontal axis represents time in minutes, and the vertical axis represents the participants, including four students (pseudonyms) and (co-)facilitators. Blue bars indicate periods of typing activity. *Phase A* (shaded blue) represents the early stage of the session, characterized by minimal and unevenly distributed typing activity. *Phase B* (shaded white) is marked by increased typing activity and shifting contributions among participants. *Phase C* (shaded pink) shows typing activity concentrated on Beto, with occasional interventions from the (co-)facilitators.

Figure 1

Generated Typing Activity Map and On-Screen Typing Detection Output



This analysis enabled us to identify critical transition points in collaboration, especially the moment when Beto assumed sole control of typing (Phase B→C), suggesting a redistribution of roles or responsibilities. A qualitative analysis of this 11.5-minute segment further illustrated this shift. At 115 minutes, just before Beto assumed full responsibility for programming, Quintanilla was actively typing code. However, she paused and asked Beto, "Do you want to type, sir? Or no?" Quintanilla wished to focus on building with Lego to create an imaginative kingdom for the final project, while Beto took over programming, modifying the code to enable the robot to take pictures of the objects (see Figure 1, right). Meanwhile, Sandrita, Emily, and Quintanilla collaborated on constructing Lego structures for the robot to detect. When Beto finished coding, the first photo taken by the robot was not ideally framed. Emily, recognizing this, stepped in to join Beto in repositioning the robot. This moment illustrates the dynamic and fluid nature of collaborative work involving multimodal forms of participation. It also highlights how students negotiated access to shared tools as roles and tasks shifted in response to evolving needs.

This study demonstrates the potential of AI-enhanced video analysis to advance the examination of multimodal participation in CSCL. Activity Maps were instrumental in revealing patterns in group interaction, including the timing, frequency, and distribution of access to the shared computer. When integrated with screen recording analysis, detected typing inputs provided meaningful context for on-screen activity. The approach offered insight not only into who contributed to programming tasks and when, but also into how evolving participation patterns were intertwined with shifts in roles and responses to emerging project demands. The study illustrates one possible use of machine-generated outputs in enriching critical human interpretation. Future work will leverage the full capabilities of the Activity Mapping system, including the detection of talking and writing activities, to provide a more comprehensive analysis of CSCL processes.

 © ISLS

References

Cukurova, M., Giannakos, M., & Martinez-Maldonado, R. (2020). The promise and challenges of multimodal learning analytics. *British Journal of Educational Technology, 51*(5), 1441-1449. https://doi.org/10.1111/bjet.13015

Engeström, Y. (1987). *Learning by expanding: An activity-theoretical approach to developmental research*. Orienta-Konsultit.

Erickson, F. (2012). Definition and analysis of data from videotape: Some research procedures and their rationales. In G. Camilli, J. Green, J. L. Green, & P. B. Elmore (Eds.), *Handbook of complementary methods in education research* (pp. 177–191). Routledge.

Jatla, V., Teeparthi, S., Egala, U., Celedón-Pattichis, S., & Pattichis, M. S. (2025). Fast and accurate video analysis and visualization of classroom activities using multiobjective optimization of extremely low-parameter models. *IEEE Access, 13*, 81143–81166. https://doi.org/10.1109/access.2025.3567437

Jordan, B., & Henderson, A. (1995). Interaction analysis: Foundations and practice. *Journal of the Learning Sciences, 4*(1), 39-103. https://doi.org/10.1207/s15327809jls0401_2

Lee, H. H. (2025). *AI-enhanced multimodal activity mapping: Understanding group dynamics through verbal, nonverbal, and object-based interactions in collaborative learning* (Doctoral dissertation). The University of Texas at Austin.

Lee, H. H., Celedón-Pattichis, S., Pattichis, M. S., Johnson, A. R., Cantú, E., Tovar, I., & LópezLeiva, C. A. (2023). *Knowing and enjoying: Expanding Latinx students' experiences with an integrated computer programming and mathematics curriculum* [Conference presentation]. Annual Meeting of the American Educational Research Association, Chicago, IL, United States. https://doi.org/10.3102/2012330

Lee, H. H., Celedón-Pattichis, S., Pattichis, M. S., LópezLeiva, C. A., & Cantú, E. (2024). I*ntegrating computer science and mathematics for enhanced learning using object-oriented programming and robotics: Mixed methods* [Conference presentation]. Annual Meeting of the American Educational Research Association, Philadelphia, PA, United States. https://doi.org/10.3102/2110022

Lee, H. H., Sung, H., Celedón-Pattichis, S., & Pattichis, M. S. (2026). Toward an inclusive understanding of collaborative learning using MMLA: Exploring multimodal participation dynamics by gender and linguistic diversity. *Learning and Instruction, 102*, 102317. https://doi.org/10.1016/j.learninstruc.2026.102317

Martinez-Maldonado, R., Gašević, D., Echeverria, V., Nieto, G. F., Swiecki, Z., & Shum, S. B. (2021). What Do You Mean by Collaboration Analytics? A Conceptual Model. *Journal of Learning Analytics, 8*(1), 126-153. https://doi.org/10.18608/jla.2021.7227

Miller, A. L., Nyegenye, S. N., & Mostafa-Shoukry, F. R. (2022). Disrupting dominant modes of expression: Illuminating the strengths and gifts of two disabled girls of color. In S. A. Annamma et al. (Eds.), *DisCrit expanded: Reverberations, ruptures, and inquiries* (pp. 45–61). Teachers College Press.

Ouyang, F., Xu, W., & Cukurova, M. (2023). An artificial intelligence-driven learning analytics method to examine the collaborative problem-solving process from the complex adaptive systems perspective. *International Journal of Computer-Supported Collaborative Learning, 18*(1), 39-66. https://doi.org/10.1007/s11412-023-09387-z

Sinclair, N., & de Freitas, E. (2019). Body studies in mathematics education: diverse scales of mattering. *ZDM Mathematics Education, 51,* 227-237. http://dx.doi.org/10.1007/s11858-019-01052-w

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Worsley, M., Martinez-Maldonado, R., & D'Angelo, C. (2021). A New Era in Multimodal Learning Analytics: Twelve Core Commitments to Ground and Grow MMLA. *Journal of Learning Analytics, 8*(3), 10-27. https://doi.org/10.18608/jla.2021.7361

Yan, L., Zhao, L., Gasevic, D., & Martinez-Maldonado, R. (2022, March). Scalability, sustainability, and ethicality of multimodal learning analytics. In *LAK22: 12th International Learning Analytics and Knowledge Conference* (pp. 13-23). https://doi.org/10.1145/3506860.3506862

Yeh, C., Reinholz, D. L., Lee, H. H., & Moschetti, M. (2025). Beyond verbal: A methodological approach to highlighting students' embodied participation in mathematics classroom. *Educational Researcher, 54*(2), 103-110. https://doi.org/10.3102/0013189X241310169

Acknowledgments

This work was supported by the University Graduate Continuing Fellowship at The University of Texas at Austin, awarded to the first author, and was also based on work supported by National Science Foundation (NSF) Awards 1613637 and 1949230. Opinions, findings, and recommendations expressed are those of the authors, not the NSF.

International Society of the Learning Sciences

Human–AI Learning Alignment in Situated Contexts: A Human-Centered Framework

Shi Ding, Georgia Institute of Technology, sding84@gatech.edu
Brian Magerko, Georgia Institute of Technology, magerko@gatech.edu

Abstract: As generative artificial intelligence (AI) transforms education, many existing approaches to AI in education emphasize task completion while overlooking learner agency and the sociotechnical nature of learning. This gap limits our ability to understand how AI supports learning in situated contexts. In this paper, we define human–AI learning alignment and present TEACH-AI (Trustworthy and Effective AI Classroom Heuristics) as a learner-centered set of analytical dimensions spanning interactional, learning-oriented, and sociotechnical perspectives. Through an illustrative case analysis, we discuss how this perspective can be used to analyze AI-supported learning settings. This work contributes a conceptual foundation for understanding how generative AI supports learning in situated contexts.

Introduction and related work

As generative AI (GenAI) becomes integrated into learning environments, existing approaches emphasize task completion while overlooking learner agency, interaction dynamics, and pedagogical goals. Prior work highlights the importance of contextualized, learner-centered support (Shaffer & Resnick, 1999; Papert, 1993). However, generative AI introduces challenges such as over-reliance, limited socio-emotional support, and risks such as bias and inaccuracy (Bender et al., 2021; Wang et al., 2024). At the same time, many HCI and human–AI approaches rely on general criteria (e.g., clarity, usefulness), domain-specific systems, or standardized tasks (Amershi et al., 2019; Faruk et al., 2025), limiting their ability to capture how AI behavior aligns with pedagogical goals in authentic learning contexts (Rosenfeld, 2021; Ribera et al., 2019; Puig et al., 2020). This paper asks: what characterizes effective human–AI learning alignment, and how can these insights inform the design and understanding of generative AI systems? We argue that current approaches fail to capture this alignment, particularly in situated learning contexts. To address this gap, we introduce human–AI learning alignment, which characterizes how AI-generated support aligns with learners' cognitive needs, pedagogical objectives, and interaction context in situated settings (Wang et al., 2024; Ding & Magerko, 2026). We present TEACH-AI, a set of analytical dimensions that integrate pedagogical goals and interaction dynamics to support context-aware analysis of generative AI in learning contexts. This paper makes two contributions: (a) we introduce human–AI learning alignment and describe it through TEACH-AI as analytical dimensions built in the literature from AI, HCI, and the learning sciences; (b) we show how this perspective supports context-aware, pedagogically grounded analysis of generative AI in situated learning.

Methodology

Following Arksey and O'Malley's (2005) framework for scoping reviews, we synthesize prior work to identify gaps in understanding generative AI in learning contexts. We reviewed Google Scholar and major venues (e.g., CHI, IDC, AIED) using keywords related to human-centered AI and educational AI systems. In total, we reviewed 126 sources across HCI, education, responsible AI, and generative AI from the pre-LLM era (before 2017) to the generative AI boom (2023–present). We iteratively coded the literature through thematic grouping and identified recurring concepts across sources. Disagreements in categorization were resolved through discussion and consultation with a senior faculty advisor to ensure consistency. Based on this synthesis, we organized the framework into three levels with eight components capturing human–AI learning alignment, and illustrated it through a case analysis.

TEACH-AI: An analytical approach and illustrative case study

In this section, we introduce TEACH-AI and illustrate its use through a brief case study.

Analytical components in the TEACH-AI framework

TEACH-AI conceptualizes human–AI learning alignment as a multi-level construct spanning feedback, learning, and sociotechnical dimensions. We identify eight core dimensions that characterize human–AI learning alignment in context (Table 1). Together, these components provide a structured lens for analyzing how generative AI supports learning processes and experiences across feedback, pedagogical, and sociotechnical contexts.

 © ISLS

Table 1
TEACH-AI conceptualizes eight analytical components for human–AI learning alignment across feedback, learning, and sociotechnical dimensions.

Level	Component	Analytical Perspective	Key References
AI Feedback	Interpretability for learning	Clarity of explanations in supporting learner understanding	(Rosenfeld, A. 2021; Ribera, M., et al. 2019)
	Interaction coherence	Coherence across interactions in the learning context	(Robnik-Šikonja, M., et al., 2018; Carvalho, D. V., et al., 2019)
	Iterative Learning Support	Support for Iterative improvement in progressive learning	(Wang, H. et al., 2024; Vasconcelos, H. et al., 2022)
Human Learning	Pedagogical usefulness	Pedagogically aligned assistance	(Faruk et al., 2025; Puig et al., 2020)
	Learning Exploration	Creativity, curiosity, and knowledge transfer	(Cherry & Latulipe, 2014; Shaffer & Resnick, 1999; Papert, 1993)
	Learner-centered adaptation	Personalization and context awareness	(Brusilovsky, P., et al., 2004; Amershi, S., et al., 2019)
Sociotechnical Context	Learning Context Alignment	Collaborative Learning Dynamics	(Xue, X.et al., 2025; Xu, F. et al., 2024)
	Responsible Learning Support	Ethical considerations shaping learners' experiences	(Bender et al., 2021; Wang, H. et al., 2024; Ding & Magerko, 2026, Morris, M. R. 2020)

Note: This table presents a conceptual structure of TEACH-AI as a lens for analyzing human–AI learning alignment in context; detailed operationalization is left for future work.

Illustrative case study

To illustrate this perspective, we show how TEACH-AI can be used to analyze generative AI assistants in situated learning contexts. The framework can be flexibly applied across different learning scenarios and priorities. For example, an AI assistant may guide learners step-by-step toward a solution but limit opportunities for independent learning, indicating a misalignment with pedagogical goals. This example highlights how TEACH-AI supports context-sensitive analysis of human–AI learning alignment.

Conclusion and future work

In summary, we introduce human–AI learning alignment and TEACH-AI, a framework for analyzing alignment across feedback, learning, and sociotechnical dimensions in context. Future work will focus on validating this framework in real-world settings.

 © ISLS

References

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., ... & Horvitz, E. (2019, May). Guidelines for human-AI interaction. In *Proceedings of the 2019 chi conference on human factors in computing systems* (pp. 1-13).

Arksey, H., & O'malley, L. (2005). Scoping studies: towards a methodological framework. *International journal of social research methodology*, *8*(1), 19-32.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021, March). On the dangers of stochastic parrots: Can language models be too big? 🦜. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency* (pp. 610-623).

Brusilovsky, P., Karagiannidis, C., & Sampson, D. (2004). Layered evaluation of adaptive learning systems. *International Journal of Continuing Engineering Education and Lifelong Learning*, *14*(4-5), 402-421.

Carvalho, D. V., Pereira, E. M., & Cardoso, J. S. (2019). Machine learning interpretability: A survey on methods and metrics. *Electronics*, *8*(8), 832.

Cherry, E., & Latulipe, C. (2014). Quantifying the creativity support of digital tools through the creativity support index. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *21*(4), 1-25.

Ding, S., & Magerko, B. (2026). Bridging responsible AI and AI literacy: The TEACH-RAI framework and toolkit for education, design, and research. In Proceedings of the 57th ACM Technical Symposium on Computer Science Education (SIGCSE) (Vol. 2).

Faruk, L. I. D., Pal, D., Funilkul, S., Perumal, T., & Mongkolnam, P. (2025). Introducing CASUX: A standardized scale for measuring the user experience of artificial intelligence based conversational agents. *International Journal of Human–Computer Interaction*, *41*(9), 5274-5298.

Friedman, B., Kahn, P., & Borning, A. (2002). Value sensitive design: Theory and methods. *University of Washington technical report*, *2*(8), 1-8.

Morris, M. R. (2020). AI and accessibility. *Communications of the ACM*, *63*(6), 35-37.

Papert, S. (1993). *The children's machine: Rethinking school in the age of the computer*. Basic Books, Inc..

Puig, X., Shu, T., Li, S., Wang, Z., Liao, Y. H., Tenenbaum, J. B., ... & Torralba, A. (2020). Watch-and-help: A challenge for social perception and human-ai collaboration. *arXiv preprint arXiv:2010.09890*.

Ribera, M., & Lapedriza, A. (2019, March). Can we do better explanations? A proposal of user-centered explainable AI. CEUR Workshop Proceedings.

Robnik-Šikonja, M., & Bohanec, M. (2018). Perturbation-based explanations of prediction models. In *Human and Machine Learning: Visible, Explainable, Trustworthy and Transparent* (pp. 159-175). Cham: Springer International Publishing.

Rosenfeld, A. (2021, May). Better metrics for evaluating explainable artificial intelligence. In *Proceedings of the 20th international conference on autonomous agents and multiagent systems* (pp. 45-50).

Ruiz, D., & Duenas, T. (2024). Towards inclusive ai: Developing a w3c-inspired accessibility benchmark for large language models. *Research Gate*.

Shaffer, D. W., & Resnick, M. (1999). " Thick" authenticity: New media and authentic learning. *Journal of interactive learning research*, *10*(2), 195-216.

Vasconcelos, H., Bansal, G., Fourney, A., Liao, Q. V., & Vaughan, J. W. (2022, October). Generation probabilities are not enough: Improving error highlighting for ai code suggestions. In *HCAI Workshop at NeurIPS*.

Wang, H., Zhang, A., Duy Tai, N., Sun, J., & Chua, T. S. (2024). Ali-agent: Assessing llms' alignment with human values via agent-based evaluation. *Advances in Neural Information Processing Systems*, *37*, 99040-99088.

Xu, F. F., Song, Y., Li, B., Tang, Y., Jain, K., Bao, M., ... & Neubig, G. (2024). Theagentcompany: benchmarking llm agents on consequential real world tasks. *arXiv preprint arXiv:2412.14161*.

Xue, X., Lu, Z., Huang, D., Wang, Z., Ouyang, W., & Bai, L. (2025). Comfybench: Benchmarking llm-based agents in comfyui for autonomously designing collaborative ai systems. In *Proceedings of the Computer Vision and Pattern Recognition Conference* (pp. 24614-24624).

International Society of the Learning Sciences

Calling, Collecting, and Transformation: Playful Artifacts as Pivots for Joint Activity and Game Design in a Collaborative VR Design Studio

Heliya Ahmadi, Xingxing Xie, Dylan Paré
hxa5384@psu.edu, xkx5097@psu.edu, dpare@psu.edu
The Pennsylvania State University

Abstract: This paper examines how artifacts and environmental features mediate collaborative design in immersive VR. In a six-week VR collaborative studio, six designers co-created an LGBTQ+ mini-golf game. By triangulating video recordings with reflective journals, we found a stimulus-to-transformation design process that emerged from the perceptual noticing of collectibles, evolving into shared ideation and transforming game mechanics. Situated within activity theory, the study contributes process-oriented methods for researching collaborative VR design.

Introduction and theoretical framework

Design work often begins with exploration, involving the gathering of references, observation of environments, and examination of materials that can spark inspiration and guide ideation (Chan et al., 2015). Virtual reality (VR) expands this exploratory dimension by enabling designers to step into shared, three-dimensional spaces, where they can co-experience environments and examine design ideas in real time (Xia et al., 2018). Where screen-based collaboration confines teams to mouse clicks and keyboard input, VR environments enable full-body movement, direct object manipulation, and shared spatial presence, which fundamentally transform how teams generate and evolve ideas together (Dorta et al., 2016; Freeman et al., 2022; Paré et al., 2019). These affordances position VR not only as a technical tool but as a unique medium for collaborative design inquiry. However, previous VR studies predominantly centered on learning outcomes (e.g., knowledge gains), and we have a limited understanding of how environmental features and artifacts mediate early-stage design ideation and collective sensemaking (Suthers, 2006).

To unpack the dynamics of learners' interactions with different elements in the collaborative VR space, we draw on activity theory (Engeström, 1987). Within activity theory, we conceptualize the "subject" as the individual player within the VR collaborative studio community, while "tools" are the virtual manipulable artifacts and environmental cues (e.g., virtual collectables) present in the space. In addition, the "object" refers to players' actions of collecting these virtual items. For the "division of labor", players searched for the collectibles and called out their locations to other players to direct everyone to the found item(s), thus organizing the group's activity. At the end of their VR field trip, players could exchange their collected items for rewards based on their choices, which aligns with the "rule" element in the activity system. Moreover, to understand how meaning evolves within the system, we integrate the concept of "expansive learning," where systemic contradictions drive the transformation of the activity system into new forms of practice (Engeström & Sannino, 2010). Furthermore, we draw on the concept of "pivots" (Vygotsky, 1967) to conceptualize how learners separate abstract meaning from immediate physical objects during this transformation. Guided by these lenses, we then ask (1) How do artifacts and environmental features in a collaborative VR design studio shape players' transformation of game design mechanics to a future game design context?

Methods

This case study (Yin, 2011) was embedded in a large six-week participatory design project from July to August 2025, exploring how six participants (aged 25–42; a mix of graduate students and VR professionals) collaboratively designed a U.S. LGBTQ+ mini-golf game. Across nine three-hour sessions, participants engaged in VR field trips (e.g., VRChat, Walkabout Mini Golf) to explore pre-designed virtual environments and VR prototyping tools like ShapesXR for design. Data sources included video recordings and 44 individual reflective journals documenting their learning and design trajectories. We then adopted two rounds of qualitative analytic coding (Emerson et al., 2011). First, we coded participants' journals line-by-line to identify initial codes based on participants' ideas or issues they suggested. Second, we conducted a fine-grained analysis, narrowing the codes toward categories we identified as of specific interest, i.e., environmental cuing (e.g., spatial features, game mechanics, and design idea generation (e.g., naming an idea out loud, drawing a design). Through these inductive, iterative processes, we identified recurring patterns (e.g., *stimulus*, *transformation*) that supported players' design

 © ISLS

ideas, revealing how players' interactions with artifacts and environmental features led to coordinated activity and evolving future design concepts.

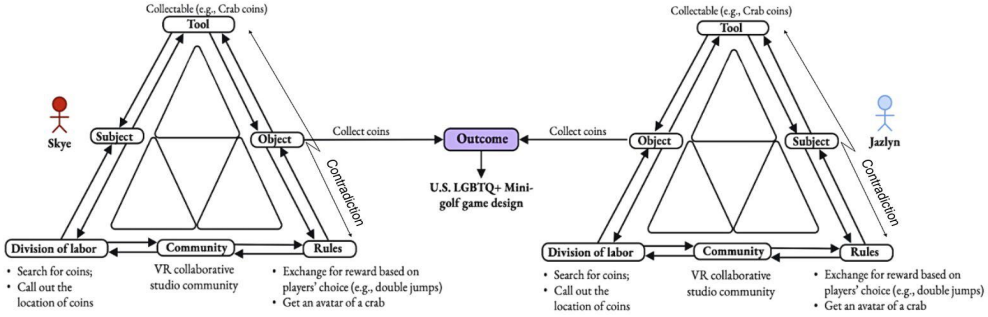
Findings: From stimulus to transformation in the coin search episode

Our analysis reveals how salient environmental features triggered an expansive learning cycle, transforming an abstract stimulus into a design mechanism. Triangulating video data with reflective journals, here, we showcase a representative episode involving two participants, Jazlyn and Skye, during a VR world exploration (i.e., *Spirit of the Sea*) (see Figure 1). For example, when Skye navigated toward a hidden collectible object (crab coins) in the VR space, she called out her discovery to direct the whole group's joint attention. This spontaneous division of labor transformed individual perception into a shared group action (collect coins). As Jazlyn corroborated this scenario in her journal, she noted that "*we all searched for glowing or hidden crab coins together*," highlighting the visual salience of the coins acted as a mediated tool.

Within this activity system, a contradiction emerged between the initial perception of the coins as mere decorative "environmental stimuli" and their "potential functional value." This contradiction drove the system's expansive transformation when the participants engaged with the mini-golf game's rules. Jazlyn reflected that the coins "*could be traded in for actions like jumping abilities* (in the game design), *which added meaning and fun to collecting them*." This game transformed the virtual objects from decorative collectibles in the environment to consequential, playful objects with exchange value. The play process of transforming an object allowed for the object to be conceptualized abstractly as a game mechanic. This builds on earlier work by Paré et al. (2019), showing how playful interactions with virtual artifacts in VR can serve as pivots for learners between lived experience and future, imagined designs. In this case, the virtual coin served as a "pivot" (Vygotsky, 1967, p.13), supporting participants to separate the abstract concept of a reward mechanism from the specific virtual object. Participants then successfully transformed this newly negotiated meaning into a cooperative progression-and-reward system for prototyping their LGBTQ+ mini-golf design in ShapeXR. This illustrates how environmental affordances and interactions in shared spaces facilitate participants' thinking with design mechanics. The coordination of joint activity in a community (calling out, finding coins) around the game rules (collect coins to gain rewards) then leads to shared meaning-making as participants reflect on this activity system to transform into their design of the mini-golf game.

Figure 1

Activity Theory Conceptualization



A diagram of the dynamics between the players (Skye and Jazlyn), their actions collecting and exchanging coins, and subsequent move toward designing the mini-golf game (outcome).

Conclusion

This study explored how artifacts and environmental features serve as design mechanisms that mediate collaborative design in immersive VR. Our findings reveal that meaning-making in VR emerges as a cyclical ecology of perception, action, and reflection, mediated by the activity system. The stimulus-to-transformation pathway aligned with the concept of expansive learning, driven by systemic contradictions regarding the utility of virtual objects. Participants did not simply consume content. They moved from noticing an environmental stimulus (a glowing coin) to ideating on the game mechanic (the exchange value of collectibles) and transforming it into a future game design context (the LGBTQ+ mini-golf game). These VR artifacts and environment designs thus also served as Vygotskian pivots that reconfigured the group's division of labor (calling out to the group) and game rules (exchange value), thereby scaffolding collective creativity in the activity system. We hope these findings contribute to learning sciences and HCI by showing how interactional features in VR environments could scaffold the shift from personal perceptual insight to shared design ideas.

 © ISLS

References

Chan, J., & Schunn, C. D. (2015). The importance of iteration in creative conceptual combination. *Cognition, 145*, 104–115.

Dorta, T., Kinayoglu, G., & Hoffmann, M. (2016). Hyve-3D and the 3D cursor: Architectural co-design with freedom in virtual reality. *International Journal of Architectural Computing, 14*, 87–102.

Emerson, R. M., Fretz, R. I., & Shaw, L. L. (2011). *Writing ethnographic fieldnotes* (2nd ed.). University of Chicago Press.

Engeström, Y., & Sannino, A. (2010). Studies of expansive learning: Foundations, findings and future challenges. *Educational Research Review*, 5(1), 1–24.

Engeström, Y. (1987). *Learning by expanding: An activity-theoretical approach to developmental research*. Orienta-konsultit.

Freeman, G., Acena, D., McNeese, N. J., & Schulenberg, K. (2022). Working together apart through embodiment: Engaging in everyday collaborative activities in social virtual reality. *Proceedings of the ACM on Human-Computer Interaction, 6*(GROUP), Article 17, 1–25.

Paré, D., Sengupta, P., Windsor, S., Craig, J., & Thompson, M. (2019). Queering virtual reality: A prolegomenon. In P. Sengupta, M-C. Shanahan, & B. Kim (Eds.), *Critical, transdisciplinary and embodied approaches in STEM education* (pp. 307–328). Springer.

Suthers, D. D. (2006). Technology affordances for intersubjective meaning making: A research agenda for CSCL. *International Journal of Computer-Supported Collaborative Learning, 1*, 315–337.

Vygotsky, L. S. (1967). Play and its role in the mental development of the child. *Soviet psychology, 5*(3), 6-18.

Xia, H., Herscher, S., Perlin, K., & Wigdor, D. (2018). Spacetime: Enabling fluid individual and collaborative editing in virtual reality. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology (UIST '18)* (pp. 853–866). Association for Computing Machinery.

Yin, R. K. (2011). *Applications of case study research* (3rd ed.). Sage Publications.

 © ISLS

International Society of the Learning Sciences

Investigating support for collaboration on the AI Frontier: A Qualitative Study of Middle-Grades Computer Science Classrooms Informed by the MOSAIC Protocol

Jeffrey Bush Tayne, University of Colorado, Boulder, jeff.bush.tayne@colorado.edu
Nga Hoang, University of Colorado, Boulder, nga.hoang-1@colorado.edu
Indrani Dey, University of Wisconsin-Madison, idey2@wisc.edu

Abstract: This study investigates the relationship between middle school student collaborative behaviors and their activities before and after receiving support while collaborating. Regression analysis of data from the MOSAIC protocol was used to conduct maximum variation sampling for a qualitative case studies to explore how student activity immediately before receiving support predicts whether they receive support focused on collaboration and if they receive this collaborative support or not predicts student activity immediately after they receive support. The characteristics of both the collaboration support provided and the nature of the collaborative tasks play crucial roles in shaping these interactions. These findings highlight the complexity of fostering effective student collaboration and underscore the need for tailored support strategies that consider teachers' pedagogical moves and the specific features of collaborative tasks.

Introduction

Collaborative learning can foster essential skills (Bower & Richards, 2006; Yoder et al., 2020), positively impact student's well-being (Johnson & Johnson, 2009), and promote equitable learning (Werner et al., 2004). However, productive collaboration, even with scripting, can be challenging and both students and teachers may need additional support for working as a group, rather than individually (Hakami, Hernández-Leo, & Amarasinghe, 2024; Fernández-Díaz et al., 2024). Simply grouping students does not automatically result in productive collaboration (Kreijins, et al., 2003). While studies have investigated support and scaffolding interactions in the classroom, most research has focused on scaffolding for individual learning (van de Pol et al., 2010), and less is known about the variety of support types in Computer-Supported Collaborative Learning (CSCL) classrooms.

Theoretical framing

Recently, researchers have started using AI-driven technologies to support collaboration, though the focus is usually on students' cognitive engagement (Ouyang & Zhang, 2024), and less on their collaboration. Additionally, understanding instructor facilitation without technological tools is essential for designing effective support systems (Holstein & Olsen, 2023). We situate our work in the context of the AI frontier to provide empirical evidence to support AI tool development's towards purposeful partnering with AI. This includes a) theoretical alignment, and b) deployment after harm-causing design issues are resolved, as can happen when AI tools obscure decision making, enable exploitation, ignore safeguards, create dependence, do not have humans in the loop, do not align with educator goals and do not attend to trust (Williamson, Molnar & Boninger, 2024; Bastani et al. 2024; Slama, Toutziaridi & Reich, 2024). Knowing what support exists for collaboration and how it is both impacted by and impacts students will guide human-centered design decisions to support teachers to have more capacity to attend to the much-needed human (and AI augmented) supports for students during collaboration.

Successful collaboration necessitates students to work together in groups to establish common ground (Roschell & Teasley, 1995), share and build on each other's ideas (Suthers et al., 2010; Scardamalia & Bereiter, 2006), negotiate ideas and coordinate the activity with each other (Barron & Roschelle, 2009), building knowledge in the process, to solve problems together. The teacher is a limited resource, especially in CSCL classrooms where they need to simultaneously facilitate and support learning at multiple levels (individual, group, and class) and coordinate across activities that often include technology, and within the constraints of time, curricula, etc. (Dillenbourg, 2013).

Methods

This study is situated within a larger study of a research-practice partnership between a large research University and educators from two rural and suburban districts in the Western United States. The data from this study comes from four middle school classrooms with a total of 302 consenting students, where the teachers were implementing a collaborative Computer Science curriculum as a part of the teachers' participation in a multi-year professional

 © ISLS

learning community. This study applies the MOSAIC protocol (Bush Tayne, Dey and Hoang, 2026; Hoang, Bush and Dey, 2024) to identify and analyze support moments, which are moments when individuals outside the observed group (e.g., teachers, other students, etc.), provide support in the form of guidance or answering questions the group may have. We identified 7 instrumental case moments based on review of the 618 support moments and their coding using the MOSAIC protocol.

Using maximum variation sampling of these seven moments, we identified two examples (from two different classrooms) for case studies to describe and explore two distinct situations, one exploratory case study came from a moment that the quantitative analysis identified was uncommon but potentially quite effective, the second case is instrumental, from a moment that the quantitative analysis identified as common and less supportive of students (Stake, 1995). In the first case, we present an exploratory case study to inform future design directions. In the second case study, we present an instrumental case to better understand common issues with this type of collaboration support).

Results

Example 1: Collaboration support modeled collaboration and is catalyzed by curriculum design features, promotes student collaboration

The first example is during the jigsaw activity, where all the group members developed expertise in different programmable sensors and then came back together to teach one another about their sensors. This example was chosen as an exploratory case study to better understand a situation that the quantitative analysis identified was uncommon. In this exploratory case, students were not collaborating before receiving support, their support focused on collaboration, and then they engaged in collaboration after the support moment.

In this example, the teacher prompted students to consider using multiple sensors and modeled how to share and integrate ideas. She also validated their contributions, indicating that their ideas were valued. These actions appear to have prompted students to begin discussing their ideas with one another, suggesting that the teacher's guidance and examples directly supported the initiation of collaborative behaviors.

Example 2: Collaboration support too surface level, no change in collaboration

In the second example, students were working on the card sort activity (described in Table 1). This moment was selected to exemplify the situation that the quantitative analysis showed was common: collaboration activity was unchanged by collaborative support, with constant characteristics before and after collaboration support. IIn this example, without additional guidance or modeling, students may not have understood what deeper collaborative behaviors entailed. Although the support was coded as collaboration-focused (as the teacher encouraged students to talk to one another), students received little instruction on what to discuss, share, or how to build on each other's ideas. As a result, they may have continued their prior behavior without engaging in more meaningful collaboration.

Conclusion

Qualitative analysis reveals nuance to MOSAIC protocol analysis of small group collaboration. Two comparative instrumental case studies analyzed the ways that the support provided relates to prior collaboration and other relevant features of the learning environment. We found that collaboration support that 1) directly modeled collaboration practices, 2) included the asking of open questions, and 3) was in the context of tasks that afforded interdependence were features of support that resulted in increased student collaboration. These findings highlight the complexity of fostering effective student collaboration and underscore the need for tailored support strategies that consider teachers' pedagogical moves and the specific features of collaborative tasks. This research contributes to the understanding of how to better attend to the human components in support to enhance collaborative learning outcomes, while AI tools work to offload teachers' cognitive demand and automate the less-human elements of supporting youth in the classroom.

References

Barron, B. & Roschelle, J. (2009). Shared cognition. In Anderman, E. (Ed.). *Psychology of Classroom Learning: An Encyclopedia*, pp. 819-823. Detroit, MI: Macmillan Reference USA.

Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R. (2024). Generative ai can harm learning. The Wharton School Research Paper.

Bower, M., & Richards, D. (2006). Collaborative learning: Some possibilities and limitations for students and teachers. *In 23rd Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education: Who's Learning* (pp. 79-89).

Bush Tayne, J., Dey, I., Hoang, N.(2026) Moments of Support Analysis in Collaboration MOSAIC: Analyzing Teacher Support During Small Group Work Video Annotation Manual. AI Institute for Student and AI teaming, University of Colorado at Boulder. https://sites.google.com/colorado.edu/mosaic-annotation-protocol/home

Fernández-Díaz, M., Daradoumis, T., Dimitriadis, Y., & Dellatola, E. (2024). "An Orchestration Perspective on Computer-Supported Collaborative Language Learning (CSCLL)." International Journal of Educational Research, 100, 102— (Sept. 2024).

Hakami, L., Hernández-Leo, D., & Amarasinghe, I. (2024). "Investigating Teacher Orchestration Load in Scripted CSCL: A Multimodal Data Analysis Perspective." British Journal of Educational Technology, 55(5), 1926-1949.

Hoang, N., Bush, J. B., & Dey, I. (2024, March). Support Patterns in Classrooms Implementing a Computer Science and Physical Computing Curriculum. In Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 2 (pp. 1674-1675).

Johnson, D. W., & Johnson, R. T. (2009). An Educational Psychology Success Story: Social Interdependence Theory and Cooperative Learning. *Educational Researcher*, 38(5), 365–379.

Ouyang, F., & Zhang, L. (2024). AI-driven learning analytics applications and tools in computer-supported collaborative learning: A systematic review. *Educational Research Review*, *44*, 100616.

Scardamalia, M., & Bereiter, C. (2006). Knowledge building: Theory, pedagogy, and technology. In *K. Sawyer (Ed.), Cambridge Handbook of the Learning Sciences* (pp. 97-118). New York: Cambridge University Press.

Slama, R., Toutziaridi, A. C., & Reich, J. (2024, July). Three Paradoxes to Reconcile to Promote Safe, Fair, and Trustworthy AI in Education. In Proceedings of the Eleventh ACM Conference on Learning@ Scale (pp. 295-299).

Stake, R. (1995). Case study research. Cham: Springer.

Suthers, D. D., Dwyer, N., Medina, R., & Vatrapu, R. (2010). A framework for conceptualizing, representing, and analyzing distributed interaction. *International Journal of Computer-Supported Collaborative Learning*, *5*, 5-42.

Werner, L. L., Hanks, B., & McDowell, C. (2004). Pair-programming helps female computer science students. *Journal on Educational Resources in Computing (JERIC)*, *4*(1), 4.

van de Pol, J., Volman, M., & Beishuizen, J. (2010). Scaffolding in teacher–student interaction: A decade of research. *Educational psychology review*, *22*, 271-296.

Williamson, B., Molnar, A., & Boninger, F. (2024). *Time for a Pause: Without Effective Public Oversight, AI in Schools Will Do More Harm Than Good*. Commercialism in Education Research Unit.

Yoder, N., Atwell, M. N., Godek, D., Dusenbury, L., Bridgeland, J. M., & Weissberg, R. (2020). *Prep aring youth for the workforce of tomorrow: Cultivating the social and emotional skills employers demand*. In SEL for Workforce Development. Collaborating States Initiative. Collaborative for Academic, Social, and Emotional Learning.

Acknowledgments

This research was supported by the NSF National AI Institute for Student-AI Teaming (iSAT) under grants DRL 2019805 and DRL 2454151. The opinions expressed are those of the authors and do not represent views of the NSF.

International Society of the Learning Sciences

Physiological Indicators of Teacher Stress in CSCL Versus Non-Collaborative Learning Sessions

Eduard Arnau Romeu, Batuhan Sayis, Davinia Hernández-Leo
eduard.arnau@upf.edu, batuhan.sayis@upf.edu, davinia.hernandez-leo@upf.edu
Universitat Pompeu Fabra

Abstract: This exploratory single-subject study examines teacher physiological stress across Computer-Supported Collaborative Learning (CSCL) versus non-CSCL sessions. One teacher was observed across 22 sessions using biosensors (heart rate variability, electrodermal activity) and questionnaires. Multivariate analysis revealed no overall physiological differences. However, exploratory analysis showed significantly elevated post-session electrodermal activity after CSCL. Given the single-subject design, findings provide methodological insights and hints for new knowledge regarding cumulative arousal, rather than broad generalizations.

Introduction

Computer-Supported Collaborative Learning (CSCL) environments create complex orchestration demands for teachers, requiring real-time monitoring of multiple student groups, technological troubleshooting, and adaptive pedagogical decisions (Dillenbourg, 2013; Ortega-Arranz et al., 2024). While research has extensively examined student outcomes in CSCL (Jeong et al., 2019), less attention has been paid to the teacher experience and physiological load during collaborative technology-mediated instruction (Amarasinghe et al., 2022).

A critical question for CSCL implementation concerns whether orchestration demands produce different physiological stress patterns compared to traditional teaching. Understanding these differences is necessary to enhance real-time, data-driven feedback and support mechanisms for educators navigating complex digital environments (Haataja et al., 2018). If stress patterns are comparable, teachers are managing demands effectively; if elevated, specific interventions are needed.

This exploratory study addresses the research question: How do teacher physiological stress indicators such as heart rate variability (HRV) and electrodermal activity (EDA) and psychological indicators from the State-Trait Anxiety Inventory (STAI) (Spielberger et al., 1971) questionnaire differ between CSCL and non-CSCL real classroom settings? CSCL sessions refer to scripted collaborative learning using technology-mediated tools, while non-CSCL sessions comprise traditional instruction with individual work and whole-class activities.

Theoretical background

Teacher orchestration in CSCL involves managing technological, pedagogical, and social dimensions simultaneously, requiring complex real-time decision-making processes (Dillenbourg et al., 2013; Ortega-Arranz et al., 2024). This multi-layered orchestration can create increased cognitive and emotional demands compared to traditional instruction (Prieto et al., 2011).

Recent studies explore teacher orchestration load using multimodal data, stressing the need for objective, continuous metrics of hidden instructional load to inform data-driven support (Hakami et al., 2024), but used only one biosignal (HRV) and could benefit from adding more data layers, such as EDA, in order to improve reliability in real classroom scenarios. Specifically, HRV reflects autonomic nervous system responses to stress and cognitive load (Bordi et al., 2018; Gedam & Paul, 2021), while EDA indicates emotional arousal related to attention and engagement (Pijeira-Díaz et al., 2018).

Comparing these physiological responses across instructional modalities can reveal whether CSCL orchestration produces distinctive stress patterns. Such insights theoretically contribute to orchestration by informing teacher preparation and support systems for sustainable CSCL implementation (Gedam & Paul, 2021).

Methodology

This study employed a single-subject design, appropriate for exploratory investigations of individual differences and within-person comparisons. The research was conducted in Universitat Pompeu Fabra, Barcelona, Spain, and received ethical approval from the institutional ethics committee.

One female teacher with over 20 years of teaching experience was observed across 22 sessions over one academic trimester. The teacher wore a biosignalsplux wearable device [PLUX Wireless Biosignals, Lisbon, Portugal] capturing electrocardiography (ECG) for HRV analysis and EDA for skin conductance measurement at 500Hz sampling rate. Sessions included CSCL sessions (n=7) using PyramidApp, a collaborative learning

 © ISLS

application that structures student group work through scripted roles, scaffolded tasks, and real-time feedback mechanisms (Manathunga et al., 2016). These sessions required the teacher to orchestrate multiple simultaneous groups, monitor technology-mediated collaboration, and provide adaptive support across teams. Traditional non-CSCL sessions (n=15) involved conventional instruction including teacher-led explanations, individual student work, whole-class discussions, and direct instruction without collaborative technology tools. Six sessions were excluded due to missing data. Leaving a final analytical sample of 16 sessions, 6 using CSCL and 10 non-CSCL.

Pre- and post-session measurements included three-minute standing baselines for physiological data and State-Trait Anxiety Inventory (STAI) assessments.

ECG signals underwent R-peak detection and artifact correction to extract inter-beat intervals, from which Root Mean Square of Successive Differences (RMSSD) was calculated using 30-second windows (Castaldo et al., 2019). EDA signals were decomposed into tonic (Skin Conductance Level, SCL) and phasic (Skin Conductance Response, SCR) components. All physiological features were normalized as percentage change from individual baseline values (reactivity %).

Independent sample comparisons between CSCL and non-CSCL sessions were conducted using Mann-Whitney U tests for individual variables (non-parametric approach given small sample size). A one-way MANOVA examined whether the combined physiological profile (RMSSD, EDA-SCL, EDA-SCR difference between pre- and post-session measurements) differed between session types. Statistical analyses were performed in IBM SPSS v31. This study acknowledges that treating repeated measures from a single participant as independent groups violates standard test assumptions. Given the exploratory single-subject nature of this study, uncorrected p-values are reported to identify potential trends rather than confirm generalizable hypotheses.

Results

A mixed-design ANOVA was conducted to examine whether the psychological and physiological changes (post minus pre session values) differed between the collaborative (CSCL) and non-collaborative (non-CSCL) groups. The analysis revealed no significant between-subjects effect of the collaboration condition ($F(1, 13)=0.044$, $p=0.838$, $\eta^2=0.003$). There was no significant interaction between the type of measure and the collaboration condition ($F(3,39)=1.072$, $p=0.372$, $\eta^2=0.076$), indicating that the specific pattern of physiological and psychological changes did not significantly differ during CSCL versus non-CSCL sessions. Multivariate tests confirmed this lack of significant interaction (Pillai's Trace=0.257, $F(3,11)=1.271$, $p=0.332$, $\eta^2=0.257$)

Mann-Whitney U tests revealed significant differences in EDA between CSCL and non-CSCL sessions. CSCL sessions showed significantly higher skin conductance in post-session measurements ($U=11.0$, $p=0.031$, $r=0.67$), with a large effect size, indicating substantially elevated sympathetic activation after a CSCL session.

Discussion and conclusion

The primary finding is that the multivariate physiological profile did not differ significantly between CSCL and non-CSCL sessions (Pillai's Trace=0.253, $p=0.576$). When HRV, EDA and STAI are considered together, CSCL orchestration does not produce a distinctly different physiological response pattern compared to traditional instruction (Prieto et al., 2011; Amarasinghe et al., 2022). As a single-subject design, these findings cannot be generalized to imply equivalence between settings or that experienced teachers universally manage demands without stress. Instead, this study provides methodological insights and hints for new knowledge regarding hidden instructional loads. This suggests teachers may effectively regulate CSCL demands without substantially elevated overall stress responses, supporting sustainable CSCL implementation while aligning with previous CSCL orchestration findings (Hakami et al., 2024), this must be interpreted cautiously given the non-significant multivariate test and requires replication.

Given the exploratory nature and small sample size (n=16), we examined individual variables. Post-session electrodermal activity was significantly elevated following CSCL sessions ($U=11.0$, $p=0.031$, $r=0.67$), with no pre-session differences ($p=0.092$). The absence of overall differences indicates CSCL does not generate substantially different demands than traditional instruction. While the pre-post design cannot assess moment-to-moment acute stress spikes, comparable pre-session baselines suggest teachers avoid sustained elevated arousal during teaching. However, elevated post-session EDA hits cumulative arousal effects showing after instruction.

Limitations include single-teacher observation, small sample size, and lack of in-session continuous monitoring. Future research should replicate across multiple teachers, include continuous physiological monitoring during sessions.

This study provides initial evidence that CSCL may not be inherently more physiologically demanding for experienced teachers, while exploratory findings suggest potential cumulative arousal patterns warranting further investigation.

 © ISLS

References

Amarasinghe, I., van Leeuwen, A., Dimitriadis, Y., Martínez-Monés, A., Hernández-Leo, D., Ulrich Hoppe, H., Wiley, K., & Martinez-Maldonado, R. (2022). Teacher orchestration load: What is it and how can we lower the burden? Rapid Community Report Series. Digital Promise and the ISLS.

Bordi, L., Okkonen, J., Mäkiniemi, J. P., & Heikkilä-Tammi, K. (2018). Communication in the digital work environment: implications for wellbeing at work. Nordic Journal of Working Life Studies, 8.

Castaldo, R., Montesinos, L., Melillo, P., James, C., & Pecchia, L. (2019). Ultra-short term HRV features as surrogates of short term HRV: A case study on mental stress detection in real life. BMC Medical Informatics and Decision Making, 19(1), 12.

Coburn, C. E., & Penuel, W. R. (2016). Research–practice partnerships in education: Outcomes, dynamics, and open questions. Educational Researcher, 45(1), 48-54.

Dillenbourg, P. (2013). Design for classroom orchestration. *Computers & Education*, 69, 485-492.

Gedam, S., & Paul, S. (2021). A review on mental stress detection using wearable sensors and machine learning techniques. IEEE Access, 9, 84045-84066.

Haataja, E., Malmberg, J., & Järvelä, S. (2018). Monitoring in collaborative learning: Co-occurrence of observed behavior and physiological synchrony explored. Computers in Human Behavior, 87, 337-347.

Hakami, L., Sayis, B., Amarasinghe, I., & Hernandez-Leo, D. (2024). Exploring teacher orchestration load in scripted CSCL: The role of heart rate variability. In Proceedings of the 17th International Conference on Computer-Supported Collaborative Learning-CSCL 2024, pp. 163-170. International Society of the Learning Sciences.

Järvelä, S., Malmberg, J., & Koivuniemi, M. (2016). Recognizing socially shared regulation by using the temporal sequences of online chat and logs in CSCL. Learning and Instruction, 42, 1-11.

Jeong, H., Hmelo-Silver, C. E., & Jo, K. (2019). Ten years of computer-supported collaborative learning: A meta-analysis of CSCL in STEM education during 2005–2014. Educational research review, 28, 100284.

Manathunga, K., & Hernández-Leo, D. (2016, September). PyramidApp: scalable method enabling collaboration in the classroom. In European conference on technology enhanced learning (pp. 422-427). Cham: Springer International Publishing.

Ortega-Arranz, A., Amarasinghe, I., Martínez-Monés, A., Asensio-Pérez, J. I., Dimitriadis, Y., Corrales-Astorgano, M., & Hernández-Leo, D. (2024). Collaborative activities in hybrid learning environments: Exploring teacher orchestration load and students' perceptions. Computers & Education, 219, 105105.

Penuel, W. R., Allen, A. R., Coburn, C. E., & Farrell, C. (2015). Conceptualizing research–practice partnerships as joint work at boundaries. Journal of Education for Students Placed at Risk (JESPAR), 20(1-2), 182-197.

Pijeira-Díaz, H. J., Drachsler, H., Kirschner, P. A., & Järvelä, S. (2018). Profiling sympathetic arousal in a physics course: How active are students?. Journal of Computer Assisted Learning, 34(4), 397-408.

Prieto, L. P., Holenko Dlab, M., Gutiérrez, I., Abdulwahed, M., & Balid, W. (2011). Orchestrating technology enhanced learning: a literature review and a conceptual framework. International Journal of Technology Enhanced Learning, 3(6), 583-598.

Spielberger, C. D., Gonzalez-Reigosa, F., Martinez-Urrutia, A., Natalicio, L. F., & Natalicio, D. S. (1971). The state-trait anxiety inventory. Revista Interamericana de Psicologia/Interamerican journal of psychology, 5(3 & 4).

Spikol, D., Ruffaldi, E., Dabisias, G., & Cukurova, M. (2018). Supervised machine learning in multimodal learning analytics for estimating success in project-based learning. Journal of Computer Assisted Learning, 34(4), 366-377.

Acknowledgements

This work has been supported by PID2023-146692OB-C33 and 2023-26 CEX2021-001195-M granted by MICIU/AEI/10.13039/501100011033, and PRCAT05825 - SDC007/25/0001000 NextGeneration EU. D. Hernández-Leo (Serra Húnter) also acknowledges the support of ICREA through the ICREA Academia programme.

International Society of the Learning Sciences

Beyond Teacher-AI Dyads: A Triadic Synergy Model for Transforming Teacher Professional Development

Kexin Li, Gaowei Chen
kxmichelle@connect.hku.hk, gwchen@hku.hk
Faculty of Education, The University of Hong Kong

Abstract: Traditional expert-led teacher professional development models often struggle with personalization, scalability, and sustainability. While AI offers new possibilities, the prevailing AI-enhanced models often rely on teacher–AI dyads, raising concerns about opacity and reduced teacher agency. Drawing upon Connectivism and Hybrid Intelligence, we propose the Triadic Synergy Model. This framework formalizes a three-way, networked, and synergistic relationship that strategically integrates the teacher's agency, AI's analytical power, and the human expert's contextual wisdom.

Introduction

Effective teaching remains one of the most important school-based influences on student learning, and high-quality teacher professional development (TPD) is therefore central to educational improvement (Hattie, 2023; Kennedy, 2016). Yet traditional TPD models, such as expert-led workshops and seminars, often struggle to achieve personalization, scalability, and sustainability (Breault, 2013; Van Veen et al., 2012). Teachers may receive standardized input and often lack sustained and personalized support that can translate professional learning into lasting classroom change (Dzingel et al., 2025).

Against this background, AI has emerged as a promising force in TPD. AI tools can provide on-demand pedagogical support, generate personalized learning pathways, and analyze classroom data to offer immediate feedback (Tan et al., 2025). These affordances suggest that AI may help address some of the long-standing constraints of traditional TPD. However, current AI-integrated TPD models often rely on a direct teacher–AI relationship, and this dyadic structure has important limitations. For example, AI-generated feedback may be difficult to interpret or experienced by teachers as opaque and difficult to trust (Jacobs et al., 2024). At the same time, a strong emphasis on automation may position teachers as passive recipients of recommendations, thereby weakening the professional agency that effective TPD requires (Li et al., 2025). These challenges suggest that the issue is not simply whether AI should be introduced into TPD, but how it should be integrated in ways that are both easy to understand and able to preserve teacher agency.

To address this gap, this paper proposes the Triadic Synergy Model, a model that helps to move beyond teacher–AI dyads by incorporating a third participant: the human expert. Drawing on Connectivism (Siemens, 2005) and Hybrid Intelligence (Cukurova, 2025), the model conceptualizes TPD as a networked and synergistic relationship among teacher, AI, and expert.

The Triadic Synergy Model: A synthesis of teacher, AI, and expert

To resolve the tensions inherent in dyadic structures, the Triadic Synergy Model emphasizes triadic synergy, where learning capacity arises from the networked interaction among three key participants: the teacher, the AI, and the human expert. The model is theoretically grounded in Connectivism, which posits that learning in the digital age occurs through the formation of connections across a distributed network (Siemens, 2005). This ensures that knowledge transfer is not unidirectional (AI-to-teacher) but multidirectional. Furthermore, drawing on principles of Hybrid Intelligence (Cukurova, 2025), the model envisions a strategic division of labor that combines the strengths of humans and AI.

As the agentic driver, the teacher remains at the center of the professional learning process. They are not passive recipients but the active orchestrators of their professional development. These core capacities of goal-directed action and strategic self-management are similar to the principles of Self-Regulated Learning (SRL), which provides the theoretical grounding for the teacher's role in our model (Zimmerman, 2002). By actively deciding when to engage with AI analytics for specific practice opportunities and when to seek the human expert for deeper pedagogical reasoning, teachers could develop adaptive agency gradually (Tao & Gao, 2017). This centrality prevents TPD from becoming a generic exercise and fosters a deep sense of ownership.

As the scalable analyst, AI contributes what traditional expert-led TPD often struggles to provide consistently: rapid data processing, pattern recognition, and personalized support at scale. These capabilities can

 © ISLS

be operationalized in several ways to support teacher development. For example, AI captures and interprets vast streams of multimodal data (e.g., student engagement patterns, teacher discourse) to reveal insights often invisible to the human eye (Zhai et al., 2021). Building on this analysis, the AI can then generate immediate, personalized feedback for the teacher, highlighting specific moments for reflection (Demszky et al., 2024). Beyond analysis and feedback, AI's generative capacity is instrumental in creating dynamic support and practice environments (Zhang et al., 2025).

However, analytics alone are not enough. This is where the human expert plays a necessary role as the contextual scaffolder. The expert's scaffolding can be realized through several vital actions. For example, they can provide an interpretive bridge to address the AI's black-box effect. By co-interpreting the AI's output with the teacher, the expert helps transform abstract data points into a coherent pedagogical narrative that resonates with the teacher's classroom experience. Moreover, the expert models pedagogical wisdom, guiding the teacher through complex situations and facilitating reflective dialogue that requires high levels of social and emotional intelligence, a domain where AI currently falls short (Li et al., 2025). The expert can also serve as an ethical guardian, promoting responsible use of AI and maintaining a critical perspective. In essence, while the AI provides the "what" (performance data), the expert helps the teacher understand the necessary technological context, construct the "so what" (pedagogical significance), and the "now what" (actionable next steps).

Emerging research on human–AI collaboration in education points in a similar direction. For instance, Cosentino et al. (2025) found that a teacher who provided guidance, posed reflective questions, and adjusted the difficulty of AI-generated problems was crucial for optimizing the learning experience. Shifting the focus to mentorship, Huang et al. (2025) discovered a complementary relationship between graduate student mentors and generative AI. Closer to our direct context of teacher education, Yu et al. (2025) developed a Hybrid Intelligence Feedback (HIF) system that uses AI to structure and summarize peer feedback on teaching videos for pre-service teachers. Their findings showed that this AI-enhanced process led to significantly deeper and more diverse reflections. The Triadic Synergy Model builds on this insight by offering a model for how such collaboration might be structured in teacher learning.

The future of the Triadic Synergy Model

The Triadic Synergy Model offers a conceptual paradigm shift from viewing AI as a mere tool to understanding it as a participating node within a collaborative human network. To move this model from conceptualization to transformative practice, several avenues for future research could be pursued.

A primary area for future research involves the intricate design and orchestration of interaction within the triadic network. As Vaccaro et al. (2024) found that effective human–AI synergy requires appropriately allocating subtasks to whichever partner is best suited for them. While existing work from student learning provides proof-of-concept for such triadic interactions, future work is encouraged to explore the emergent dynamics of the fully integrated three-node network within teacher education. This includes developing specific protocols for orchestrating three-way interactions, creating structured activities that leverage each participant's strengths, and establishing clear coordination mechanisms for collaboration within the triadic system.

A second research direction concerns sustainability. One of the model's longer-term possibilities is a teacher-to-expert pathway, in which experienced participants gradually develop the capacity to scaffold others within the triadic network. This transition is feasible because the model synergizes the three participants to redefine and streamline the expert's responsibilities. With AI handling activity delivery and teachers driving the learning process, the primary role of experts is to facilitate dialogue and co-interpret AI-generated suggestions. This less burdensome role becomes much more manageable for experienced practitioners to adopt. Nevertheless, future studies are needed to investigate what kinds of knowledge, dispositions, and AI literacy are needed for teachers to move credibly into this expert role, and whether such near-peer experts can provide support comparable to that of instructional coaches, researchers, or master teachers.

References

Breault, D. A. (2013). The challenges of scaling-up and sustaining professional development school partnerships. *Teaching and Teacher Education*, 36, 92-100.

Cosentino, G., Anton, J., Sharma, K., Gelsomini, M., Giannakos, M., & Abrahamson, D. (2025). Hybrid teaching intelligence: Lessons learned from an embodied mathematics learning experience. *British Journal of Educational Technology*, 56(2), 621-649.

Cukurova, M. (2025). The interplay of learning, analytics and artificial intelligence in education: A vision for hybrid intelligence. *British Journal of Educational Technology*, *56*(2), 469-488.

Demszky, D., Liu, J., Hill, H. C., Jurafsky, D., & Piech, C. (2024). Can Automated Feedback Improve Teachers' Uptake of Student Ideas? Evidence From a Randomized Controlled Trial in a Large-Scale Online Course. *Educational Evaluation and Policy Analysis*, *46*(3), 483–505.

Dzingel, H., Richter, E., Klusmann, U., & Richter, D. (2025). The effects of formal and informal learning opportunities on teachers' ICT-related attitudes and self-efficacy. *European Journal of Teacher Education*, 1-19.

Hattie, J. (2023). *Visible learning: The sequel: A synthesis of over 2,100 meta-analyses relating to achievement*. Routledge.

Huang, Y., Li, S., & Liu, Z. (2025). Can GenAI complement supervisor support in shaping postgraduates' research experiences? A mixed-methods approach. *Studies in Higher Education*, 1-19.

Jacobs, J., Scornavacco, K., Clevenger, C., Suresh, A., & Sumner, T. (2024). Automated feedback on discourse moves: Teachers' perceived utility of a professional learning tool. *Educational Technology Research and Development*, *72*(3), 1307-1329.

Kennedy, M. M. (2016). How does professional development improve teaching?. *Review of Educational Research*, *86*(4), 945-980.

Li, K., Wang, P., & Chen, G. (2025). How can AI be integrated into teacher professional development programs? A systematic review based on an adapted technology-based learning model. *Teaching and Teacher Education*, *168*, 105219.

Siemens, G. (2005). Connectivism: Learning as network-creation. *ASTD Learning News*, *10*(1), 1-28.

Tan, X., Cheng, G., & Ling, M. H. (2025). Artificial intelligence in teaching and teacher professional development: A systematic review. *Computers and Education: Artificial Intelligence*, *8*, 100355.

Tao, J., & Gao, X. (2017). Teacher agency and identity commitment in curricular reform. *Teaching and Teacher Education*, *63*, 346-355.

Vaccaro, M., Almaatouq, A., & Malone, T. (2024). When combinations of humans and AI are useful: A systematic review and meta-analysis. *Nature Human Behaviour*, *8*(12), 2293-2303.

Van Veen, K., Zwart, R., & Meirink, J. (2012). What makes teacher professional development effective?: A literature review. *Teacher Learning that Matters*, 3-21.

Yu, J., Yu, S., & Chen, L. (2025). Using hybrid intelligence to enhance peer feedback for promoting teacher reflection in video-based online learning. *British Journal of Educational Technology*, *56*(2), 569-594.

Zhai, X., Chu, X., Chai, C. S., Jong, M. S. Y., Istenic, A., Spector, M., ... & Li, Y. (2021). A Review of Artificial Intelligence (AI) in Education from 2010 to 2020. *Complexity*, *2021*(1), 8812542.

Zhang, N., Ke, F., Dai, C. P., Southerland, S. A., & Yuan, X. (2025). Seeking to support preservice teachers' responsive teaching: Leveraging artificial intelligence-supported virtual simulation. *British Journal of Educational Technology*, *56*(3), 1148-1169.

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory into Practice*, 41(2), 64-70.

Acknowledgement

This work was supported by Hong Kong RGC, University Grants Committee (Grant No.: 17619325).

International Society of the Learning Sciences

The Effect of Group-Awareness on Peer Feedback Quality in Pre-Service Teachers' Collaborative Diagnostic Reasoning

Sebastian Tews, sebastian.tews@uni-hamburg.de, University of Hamburg
Melissa Özsoy, melissa.oezsoy@lmu.de, Ludwig-Maximilians-Universität München
Meral Roeben, roeben@anglistik.uni-muenchen.de, Ludwig-Maximilians-Universität München
Yasemin Erdemgil, erdemgil@anglistik.uni-muenchen.de, Ludwig-Maximilians-Universität München
Xiao Zhang, x.zhang@anglistik.uni-muenchen.de, Ludwig-Maximilians-Universität München
Christiane Lütge, luetge@anglistik.uni-muenchen.de, Ludwig-Maximilians-Universität München
Freydis Vogel, freydis.vogel@uni-hamburg.de, University of Hamburg

Abstract: This quasi-experimental study examined the effect of a group awareness tool on peer feedback quality in video-based collaborative diagnostic reasoning. In total, 54 pre-service teachers collaboratively diagnosed a teaching video in dyads and generated peer feedback for the acting teacher regarding the use of digital media in the classroom. Results suggest a rather detrimental effect on feedback quality when supported by the group awareness tool, indicating the need for additional scaffolding to utilize group awareness.

Introduction

Peer feedback is considered crucial for teachers' professional development in all phases of teacher education (Hammerness et al., 2005). According to the classification of Hattie & Timperley (2007), effective feedback comprises the definition of a clear goal, information about the progress on achieving that goal (Feed Back) and suggestions for improvement (Feed Forward). Both feedback provider and recipient are usually on a similar expertise level during peer feedback. As a form of co-constructive teacher collaboration, peer feedback holds potential for enhancing instructional quality, particularly the successful integration of digital media in the classroom (Drossel & Heldt, 2022).

Research suggests that diagnostic competences are prerequisites for delivering effective peer feedback on teaching (Prilop et al., 2024). In the teaching context, diagnosing can be understood as „the goal-oriented collection and interpretation of case-specific or problem-specific information to reduce uncertainty in order to make […] educational decisions" (Heitzmann et al., 2019, p. 4). While earlier research in teacher education predominantly focused on individual diagnosis, recent studies increasingly underscore the value of collaborative approaches (e.g. Csanadi et al., 2020). Collaborative Diagnostic Reasoning (CDR) conceptualizes the process of generating a joint diagnosis through engagement in both diagnostic and collaborative problem-solving activities (Radkowitsch et al., 2021). CSCL literature highlights that groups need instructional support to effectively engage in collaborative problem-solving (Vogel & Fischer, 2020). Research in medical education indicates that practitioners often fail to share and elicit relevant information while collaboratively diagnosing patient cases which negatively effects the diagnostic outcome (Radkowitsch et al., 2021; Brandl et al., 2024).

One concept connected to promoting collaborative learning that has received increasing attention over the past years is group awareness (GA). GA refers to the perception of relevant characteristics of the group or group members, such as their knowledge or perspectives (Bodemer & Schnaubert, 2018). GA-tools visualize such information and implicitly guide collaborative processes. One particular potential of GA-tools is facilitating the recognition of conflicting assumptions within the group. Knowledge about conflicting viewpoints can then trigger social learning activities to resolve socio-cognitive conflicts, e.g. sharing of unshared knowledge or elaborating on content-related aspects of the task by activating internal collaboration scripts (Schnaubert & Bodemer, 2022; Engelmann & Hesse, 2011). The use of a GA-tool could help address the problem that sharing and eliciting relevant information is often neglected in CDR.

Research aim and questions

CDR is a promising approach for improving teachers' professional development. However, only a few studies examined CDR in teacher education. Hence, little is known about the effectiveness of scaffolds, like GA-tools, on the diagnostic process and learning outcomes (Brandl et al., 2024). In this study the learning outcome was assessed by examining features of a written peer feedback on teaching situations, since high quality feedback can be understood as the result of the engagement in diagnostic activities (Prilop et al., 2024). The following research questions were addressed:

RQ1: What is the effect of a GA-tool during CDR on Feed Back quality?

 © ISLS

RQ2: What is the effect of a GA-tool during CDR on Feed Forward quality?

RQ3: What is the effect of a GA-tool during CDR on the feedback content?

Method

In a quasi-experimental study N = 54 teacher education students (mean age: 22.6 years; f = 36) were grouped in 27 dyads and assigned into two conditions. First, the students watched a scripted teaching video which showed the use of digital media in an English lesson. Then, they individually engaged in diagnostic activities, namely evidence generation and evidence evaluation, which involved describing and analyzing the integration of digital media in the classroom. For each analysis, participants indicated their agreement with seven items on a four-point Likert scale with each item representing an indicator of digitally enhanced teaching based on the European Framework for the Digital Competence of Educators (Redecker, 2017; Özsoy et al., 2024). Subsequently, participants were connected via audio-chat and they collaboratively developed suggestions for improvement for the observed teacher based on their individual evaluations. In the control group, participants could only see their own evaluation of the indicator of digitally enhanced teaching during collaboration. In the experimental group, participants could access both their own and their learning partners' evaluation in an assessment grid revealing agreement or disagreement for each indicator. Subsequently, participants in both conditions individually formulated a written peer feedback for the observed teacher.

A coding scheme was developed capturing the quality of the feedback messages based on the classification of Hattie & Timperley (2007) and the criteria of the Feedback Quality Index by Prins et al. (2006). Accordingly, we distinguished between the evaluation of the observed teacher behavior, the Feed Back, and suggestions for improvement, the Feed Forward. As quality criteria we chose the level of detail for specificity and explanation for both the Feed Back and Feed Forward. Additionally, we operationalized the feedback content by the level of integration of the seven indicators of digitally enhanced teaching resulting in five categories (Feed Back Specificity, Feed Back Explanation, Feed Forward Specify, Feed Forward Explanation and Use of Indicators). Our coding scheme distinguished between three quality levels (0-2). Two raters coded half of the feedback message independently, taking the whole feedback message as unit of analysis. Discrepancies between the two raters were negotiated until a consensus was reached. The interclass correlation coefficient ranged from .69 and .84 for each category before the negotiations for consensus.

Results and discussion

Descriptive results revealed that the control group achieved higher mean values in three categories. Regarding Feed Back specificity, students who received no GA support described situations in more detail on average. Conversely, the experimental group demonstrated more depth in their Feed Back explanation. In terms of Feed Forward, the control group outperformed the experimental group in both their level of detail for specificity and explanation. Concerning feedback content, the experimental group referred more strongly to the seven indicators of digitally enhanced teaching. However, partly medium effect sizes were found but none were statistically significant (all $t(52) < 1.4$, $p > .09$, $d < .36$) which may be due to the small sample size. The specific use of the GA-tool, which showed (dis)agreement concerning the evaluation of instructional features, might not be sufficient to foster collaborative diagnostic activities, as the rather superficial revelation of agreement might have led to an increase in quick consensus building (Weinberger & Fischer, 2006). The analysis of the recorded conversations will shed light on the collaborative mechanisms that influence the quality and the content of the peer feedback and will inform additional need for support during teachers' CDR.

References

Bodemer, D., & Schnaubert, L. (2020). Group Awareness-Tools beim technologieunterstützen Lernen. In Springer eBooks (pp. 321–331). https://doi.org/10.1007/978-3-662-54368-9_30

Brandl, L., Stadler, M., Richters, C., Radkowitsch, A., Fischer, M. R., Schmidmaier, R., & Fischer, F. (2024). Collaborative problem-solving in knowledge-rich domains: A multi-study structural equation model. International Journal of Computer-Supported Collaborative Learning, 19(3), 341–368. https://doi.org/10.1007/s11412-024-09425-4

Csanadi, A., Kollar, I., & Fischer, F. (2020). Pre-service teachers' evidence-based reasoning during pedagogical problem-solving: better together? European Journal of Psychology of Education, 36(1), 147–168. https://doi.org/10.1007/s10212-020-00467-4

Drossel, K., & Heldt, M. (2022). Teaching-Related Teacher Collaboration in the Age of Digital Transformation. MedienPädagogik Zeitschrift Für Theorie Und Praxis Der Medienbildung, 49, 68–93. https://doi.org/10.21240/mpaed/49/2022.06.23.x

 © ISLS

Engelmann, T., & Hesse, F. W. (2011). Fostering sharing of unshared knowledge by having access to the collaborators' meta-knowledge structures. Computers in Human Behavior, 27(6), 2078–2087. https://doi.org/10.1016/j.chb.2011.06.002

Hammerness, K. M., Darling-Hammond, L., Bransford, J., Berliner, D. C., Cochran-Smith, M., McDonald, M., et al. (2005). How teachers learn and develop. In L. Darling-Hammond, J. Bransford, P. LePage, K. Hammerness, & H. Duffy (Eds.), *Preparing teachers for a changing world: What teachers should learn and be able to do* (pp. 358-389). San Francisco, CA: Jossey-Bass.

Hattie, J., & Timperley, H. (2007). The power of feedback. Review of Educational Research, 77(1), 81–112. https://doi.org/10.3102/003465430298487

Heitzman, N., Seidel, T., Opitz, A., Hetmanek, A., Wecker, C., Fischer, M., Ufer, S., Schmidmaier, R., Neuhaus, B., Siebeck, M., Stürmer, K., Obersteiner, A., Reiss, K., Girwidz, R. & Fischer, F. (2019). Facilitating Diagnostic Competences in Simulations: A Conceptual Framework and a Research Agenda for Medical and Teacher Education. Frontline Learning Research, 7(4), 1–24. http://files.eric.ed.gov/fulltext/EJ1232563.pdf

Oezsoy, M., Murbök, J., Schultz-Pernice, F., Gräsel, C., Sailer, M. und Fischer, F. (2024). Lehrkräftekooperation im Kontext digitaler Schulentwicklung - Kokon: DigCompEduObserve – Ein Beobachtungsinstrument zur Förderung digitaler Kompetenzen von Lehrenden. Munich Center of the Learning Sciences: MCLS Reports, Nr. 3

Prilop, C., Weber, K., Rotsaert, T., & Keppens, K. (2024). Knowledge, self-efficacy and professional vision: how do they predict pre-service teachers' peer feedback quality? In A. Gegenfurtner & R. Stahnke (Eds.), *Teacher professional vision: empirical perspectives*. Routledge.

Prins, F. J., Sluijsmans, D. M. A., & Kirschner, P. A. (2006). Feedback for general practitioners in training: quality, styles, and preferences. Advances in Health Sciences Education, 11(3), 289–303. https://doi.org/10.1007/s10459-005-3250-z

Radkowitsch, A., Sailer, M., Schmidmaier, R., Fischer, M. R. & Fischer, F. (2021). Learning to diagnose collaboratively – Effects of adaptive collaboration scripts in agent-based medical simulations. Learning And Instruction, 75, 101487. https://doi.org/10.1016/j.learninstruc.2021.101487

Redecker, C. European Framework for the Digital Competence of Educators: DigCompEdu. Punie, Y. (ed). EUR 28775 EN. Publications Office of the European Union, Luxembourg, 2017, ISBN 978-92-79-73494-6, doi:10.2760/159770, JRC107466

Schnaubert, L., & Bodemer, D. (2022). Group awareness and regulation in computer-supported collaborative learning. International Journal of Computer-Supported Collaborative Learning, 17(1), 11–38. https://doi.org/10.1007/s11412-022-09361-1

Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. Computers & Education, 46(1), 71–95. https://doi.org/10.1016/j.compedu.2005.04.003

Vogel, F. & Fischer, F. (2020). Computerunterstütztes kollaboratives Lernen. In: Niegemann, H., Weinberger, A. (Eds.) *Handbuch Bildungstechnologie*. Springer. https://doi.org/10.1007/978-3-662-54368-9_3

International Society of the Learning Sciences

Comparing High School Students' Collaborative Discourse in Idea-Centered and Procedure-Based Technology-Supported Scientific PBL

Ying-Tien Wu, Ti-Cheng Tsai, Li-Jen Wang
ytwu@cl.ncu.edu.tw, 113584004@cc.ncu.edu.tw, tommywang@ncu.edu.tw
National Central University, Taiwan

Abstract: This quasi-experimental study compared collaborative discourse between idea-centered and procedure-based science PBL among 61 Taiwanese high school students over 12-week. Using a hierarchical coding framework, student contributions were coded into three levels—information sharing, knowledge construction, and knowledge building—at individual and group levels. Results showed idea-centered PBL significantly increased knowledge-building discourse (73.6% vs. 34%) and promoted idea diversification and synthesis. These findings suggest that epistemological framing embedded in platform design shapes students' collaborative discourse.

Introduction

Collaborative discourse, through which high school students exchange and build on each other's ideas to advance shared understanding, is fundamental to science project-based learning (PBL) (Scardamalia & Bereiter, 2014; Hmelo-Silver, 2004). Through such discourse in science PBL, students have the opportunities to negotiate meaning and co-construct explanation, mirroring authentic scientific practice (Scardamalia & Bereiter, 2014). However, traditionally, the implementation of science PBL has been largely procedure-based, with students following structured steps rather than engaging in authentic collaborative inquiry, thereby limiting students' collaborative discourse as participants in a scientific community (Pedaste et al., 2015).

In recent years, Knowledge Building (KB) theory, an idea-centered approach reframing students as members of a knowledge-building community and highlighting the importance of agency and collaborative discourse, has been proposed (Scardamalia & Bereiter, 2014). To promote high school students' collaborative discourse, KB is promising for designing and implementing science PBL. However, little empirical research has been conducted with this perspective. Therefore, the current study aims to address this research gap by comparing two groups of high school students' collaborative discourse during two different science PBLs, idea-centered and procedure-based. The research questions of this study are: RQ1: What types of collaborative discourse do students have in idea-centered and procedure-based science PBLs? RQ2: Are there significant differences in the distribution of collaborative discourse types between the two groups of students having idea-centered and procedure-based science PBLs?

Methods

With a quasi-experimental design, this study was conducted with two classes of second-year students (N = 61) at a high school in southern Taiwan. One class was assigned to the Idea-centered Group (n = 30), in which an idea-centered science PBL was implemented, while the other group was assigned to the Procedure-based PBL Group (n = 31), in which a conventional procedure-based science PBL was implemented. The two groups of students were asked to conduct science projects in groups (four or five students per group) for 12 weeks, a 2-period-per-week investigation into "maximum load capacity." The science PBL task required students to collaboratively design and test solutions to determine the maximum load capacity of a given structure, involving hypothesis generation, data collection, and iterative refinement of explanations.

To promote students' collaborative discourse, an idea-centered platform, in which students could generate and refine their ideas, for example, students in the idea-centered condition were prompted with open-ended questions such as "What ideas can explain this phenomenon?" and "Can you build on or challenge a peer's idea?", whereas students in the procedure-based group followed procedure-based instructions such as "Record your observations" and "Complete the worksheet." The platform guided students without explicitly requiring specific discourse types were developed in this study. Critically, neither platform explicitly instructed students to produce knowledge-building discourse; they differed in whether they afforded open-ended idea exploration or step-by-step task completion—reflecting the distinction between problematizing and structuring functions of scaffolding (Reiser, 2004).

 © ISLS

An adapted hierarchical framework (Yang et al., 2020; Hong & Sullivan, 2009; Van Aalst, 2009) was used to code contributions into three levels: Information Sharing, Knowledge Construction, and Knowledge Building (including diversification, elaboration, and synthesis). For RQ1, individual student contributions were the unit of analysis (idea-centered group: N = 364; procedure-based group: N = 400); for RQ2, groups were analyzed (8 per class; total N = 16). Chi-square tests addressed RQ1, and Mann–Whitney U tests addressed RQ2.

Results

RQ1: Students' collaborative discourse types in two different Scientific PBL activities

Table 1 presents discourse type distributions. Chi-square results indicate significant differences ($\chi^2 = 126.46$, $p < .001$): Idea-centered students produced 73.6% knowledge-building discourse vs. 34% in the procedure-based students. Procedure-based students produced substantially more information-sharing discourse (44.5% vs. 13.5%).

Table 1. Distribution of Discourse Types across Instructional Conditions

Discourse Type	Idea-centered scientific PBL (n=30)	Procedure-based scientific PBL (n=31)
	N (%)	N (%)
Information sharing	49 (13.5%)	178 (44.5%)
Knowledge construction	47 (12.9%)	86 (21.5%)
Knowledge building	268 (73.6%)	136 (34%)
χ^2	126.46***	

*** $p < .001$

RQ2: Differences in collaborative discourse types between idea-centered and procedure-based scientific PBL activities

Mann-Whitney U tests comparing the two groups (n = 8 classes each) revealed significant group differences in idea-diversification (Idea-centered Group: M = 26.25, S.D. = 9.00 vs. Procedure-based Group: M = 11.38, S.D. = 3.50; U = 60.5, p = .003, r = 0.74) and idea-synthesis (Idea-centered Group: M = 2.00, S.D. = 1.93 vs. Procedure-based Group: M = 0.25, S.D. = 0.71; U = 49.5, p = .039, r = 0.51), with large and medium effect sizes respectively. Idea-elaboration showed no significant difference (Idea-centered Group: M = 5.25, S.D. = 5.23 vs. Procedure-based Group: M = 5.38, S.D. = 3.29; U = 28.5, p = .751, r = 0.08). Particularly, the idea-centered group showed considerably higher variability in idea-diversification (S.D. = 9.00) compared to the Procedure-based group (S.D. = 3.50).

Discussion, limitations and conclusion

Idea-centered scientific PBL was associated with substantially higher levels of knowledge-building discourse (73.6% vs. 34%), suggesting that idea-centered platform design is associated with more generative forms of student participation (Tong & Chan, 2023; Scardamalia & Bereiter, 2006; Stahl & Hakkarainen, 2021). When pedagogies foreground ideas rather than procedural-based tasks, students shift from reporting information to building knowledge—consistent with KB theory's emphasis on idea agency and community knowledge advancement (Scardamalia & Bereiter, 2014).The subcategory analysis further indicates that idea diversification and synthesis may emerge within a relatively short time under idea-centered conditions, whereas idea elaboration appears to require longer-term engagement (Zhang et al., 2011). The higher variability in idea-diversification within the idea-centered group suggests that group dynamics may play a key role in shaping how idea-centered platform affordances translate into collaborative discourse. This study has several limitations. The quasi-experimental design and modest sample size limit generalizability. The 12-week duration might not be enough to capture the development of deeper epistemic processes such as elaboration. Furthermore, the findings are based on discourse patterns rather than direct measures of learning outcomes. Future studies should explore the connection between student discourse and their actual understanding. Specifically, researchers might adjust platform features to see exactly how they affect the results. Overall, these findings suggest that the way a platform frames learning—as idea improvement versus task completion—directly shapes what students say and how they say it. Platform design is not a background condition; it is a pedagogical choice.

 © ISLS

References

Hmelo-Silver, C. E. (2004). Problem-Based Learning: What and How Do Students Learn? *Educational Psychology Review*, *16*, 235-266. https://doi.org/10.1023/B:EDPR.0000034022.16470.f3

Hong, H.-Y., & Sullivan, F. R. (2009). Towards an idea-centered, principle-based design approach to support learning as knowledge creation. *Educational Technology Research and Development*, *57*(5), 613-627. https://doi.org/10.1007/s11423-009-9122-0

Pedaste, M., Mäeots, M., Siiman, L. A., De Jong, T., Van Riesen, S. A., Kamp, E. T., Manoli, C. C., Zacharia, Z. C., & Tsourlidaki, E. (2015). Phases of inquiry-based learning: Definitions and the inquiry cycle. *Educational Research Review*, *14*, 47-61. https://doi.org/10.1016/j.edurev.2015.02.003

Reiser, B. J. (2004). Scaffolding Complex Learning: The Mechanisms of Structuring and Problematizing Student Work. *Journal of the Learning Sciences*, *13*(3), 273-304. https://doi.org/10.1207/s15327809jls1303_2

Scardamalia, M., & Bereiter, C. (2006). Knowledge building. *The Cambridge*.

Scardamalia, M., & Bereiter, C. (2014). Knowledge Building and Knowledge Creation: Theory, Pedagogy, and Technology. In R. K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (2 ed., pp. 397-417). Cambridge University Press. https://doi.org/10.1017/CBO9781139519526.025

Stahl, G., & Hakkarainen, K. (2021). Theories of CSCL. In *International handbook of computer-supported collaborative learning* (pp. 23-43). Springer. https://doi.org/10.1007/978-3-030-65291-3_2

Tong, Y., & Chan, C. K. K. (2023). Promoting knowledge building through meta-discourse and epistemic discourse understanding. *International Journal of Computer-Supported Collaborative Learning*, *18*(3), 353-391. https://doi.org/10.1007/s11412-023-09403-2

Van Aalst, J. (2009). Distinguishing knowledge-sharing, knowledge-construction, and knowledge-creation discourses. *International Journal of Computer-Supported Collaborative Learning*, *4*(3), 259-287. https://doi.org/10.1007/s11412-009-9069-5

Yang, Y., van Aalst, J., & Chan, C. K. (2020). Dynamics of reflective assessment and knowledge building for academically low-achieving students. *American educational research journal*, *57*(3), 1241-1289. https://doi.org/10.3102/0002831219872444

Zhang, J., Hong, H.-Y., Scardamalia, M., Teo, C. L., & Morley, E. A. (2011). Sustaining knowledge building as a principle-based innovation at an elementary school. *The Journal of the Learning Sciences*, *20*(2), 262-307. https://doi.org/10.1080/10508406.2011.528317

Acknowledgments

This research was supported by the National Science and Technology Council (NSTC), Taiwan (112-2410-H-008 -036 -MY3, 114-2423-H-008-002). The authors gratefully acknowledge the financial support provided.

International Society of the Learning Sciences

EnactingBoard: AI-mediated Dialogue for Developing Empathic Negotiation Skills in Higher Education

Hideyuki Suzuki, Ibaraki University, hideyuki@suzuki-lab.net
Hideo Funaoi, Soka University, funaoi@umegumi.net
Yoshihiko Kubota, Tamagawa University, kubota@kubota-lab.net
Hiroshi Kato, The Open University of Japan, hkato@ouj.ac.jp

Abstract: This study proposed a three-step learning method to cultivate the skill of sustaining dialogue with emotional opponents and implemented it in the form of the EnactingBoard web application. This system integrates generative AI to simulate adversarial dialogue, reflection, empathic understanding, and counterfactual analysis. This study demonstrates how AI-mediated dialogue can provide a novel approach for developing empathic negotiation skills in higher education.

Introduction

In contemporary society, which increasingly emphasizes diversity and inclusion, individuals must collaborate with others who hold different values and perspectives. In such contexts, the ability to express opinions, persuade others, and build consensus is essential, and higher education is expected to cultivate these abilities (UNESCO, 2020). A major obstacle to consensus-building lies not in logic or evidence but in engaging with emotional opponents, who may reject arguments regardless of reasoning. This can lead students to view such opponents as "monsters" or incomprehensible beings, causing dialogue to collapse. Therefore, the ability to sustain dialogue and pursue consensus with emotional opponents is crucial for university students. However, training in this area has largely been limited to professional conflict resolution (cf., Winslade et al., 1998) and rarely appears in general higher education. This study proposes a learning method supported by an AI-based tool to foster empathic understanding of emotional opponents as a foundation for sustaining dialogue under emotional tension and disagreement.

Proposed method

We propose the following methods to operationalize three elements. (1) *Safe and realistic dialogue practice with emotional opponents*: Learners engage in adversarial dialogue with a generative AI acting as an emotional opponent whose persona is defined by the learner. Compared with peer role-play, generative AI can produce more unexpected responses and safely evoke stronger emotions. (2) *Scaffolding for empathic understanding of emotional opponents*: This process unfolds in three stages. First, learners analyze the emotions and background circumstances underlying both parties' utterances in an adversarial dialogue. Next, they explore past experiences that may underlie the opponent's hostility while also reflecting on their own past experiences that led them to adopt views opposed to those of the emotional opponents, thereby considering possible similarities between themselves and the opponent. This design draws on prior research demonstrating the role of perceived similarity between the self and others in fostering empathic understanding (Gerace et al., 2015; Hoffman, 2000; Nickerson, 1999). Finally, based on the understanding developed through these analyses, learners re-engage in dialogue to test the adequacy of their interpretation. (3) *Expansion of dialogue possibilities for sustained dialogue*: Sustaining dialogue requires not only empathic understanding of emotional opponents but also the ability to anticipate a range of possible dialogue trajectories. To support such reflection, the system generates two counterfactual dialogues, one leading to continued dialogue and the other to rupture, allowing learners to compare actual and possible paths toward sustained dialogue (List, 2006; Morris et al., 2000).

Preliminary investigation

The proposed method was implemented with 18 university students. Each student identified an opponent, engaged in dialogue, and analyzed the interaction. Generative AI was used for the dialogue and presentation of dialogue trajectories. To examine empathy toward the opponent, students rated the perceived distance between themselves and the opponent using the Inclusion of Other in the Self Scale (Aron et al., 1992) at four time points: before the dialogue task (Step 0), after Step 1, after Step 2, and after Step 3. The mean ratings were 5.40 (0.43), 6.06 (0.33), 3.80 (0.46), and 3.73 (0.39), respectively, with higher values indicating greater distance. Analysis of variance showed a significant effect of step ($F(3, 42) = 12.13$, $p < .001$). Holm's multiple comparison test showed that the perceived distance increased after the dialogue experience and then decreased after Step 2. No significant change

 © ISLS

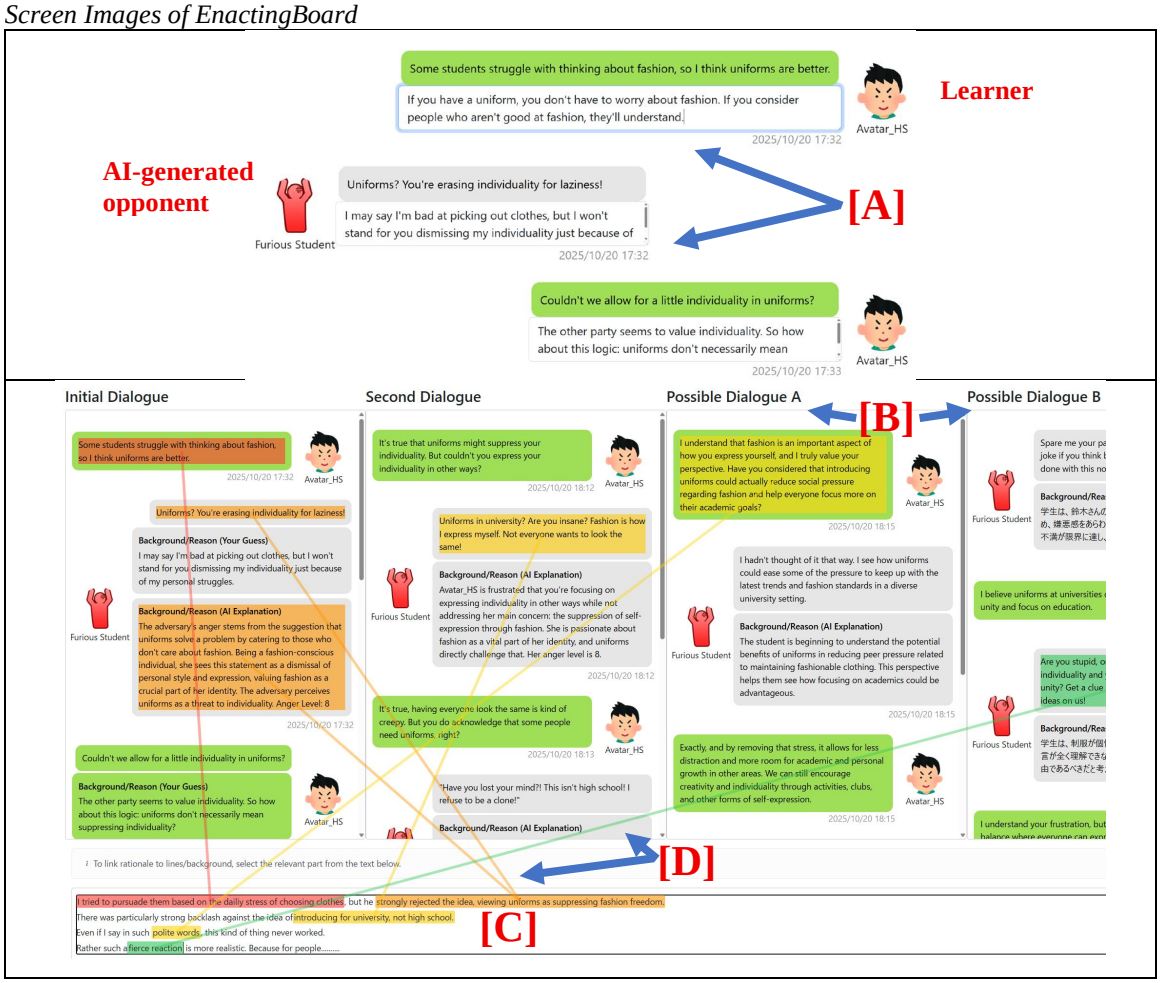
was found between Steps 2 and 3. These results suggest that the AI-based dialogue provided a realistic sense of adversarial tension. Furthermore, dialogue analysis and analysis of similarities between the self and the opponent may enhance empathy toward the opponent. Counterfactual reflection did not further increase empathy, possibly because this step served a different function, namely, expanding awareness of dialogue possibilities, although further investigation is needed to confirm this interpretation. This implementation also suggested the importance of reducing learners' procedural load for further evaluation of the proposed method.

System development: EnactingBoard

To support the implementation of the proposed method, we developed EnactingBoard (Figure 1), a web-based application built on the ChatGPT API. The system provides five modules. (1) *Dialogue*: Learners define an emotional opponent's persona and engage in dialogue through a WhatsApp-style interface. (2) *Behind the Lines*: The system presents the dialogue transcript for annotation of the emotions and circumstances underlying each utterance (Figure 1 [A]). (3) *Persona Exploration*: The system displays the opponent's persona and an AI-generated estimate of the learner's persona from the opponent's perspective, supporting reflection on past experiences and possible similarities. (4) *Re-dialogue*: The system generates a new dialogue session with a hidden emotional key, which learners attempt to identify through continued dialogue. (5) *Counterfactual Reflection*: The system juxtaposes the learner's dialogue with two AI-generated alternatives—an ideal end and a disastrous end—while supporting written reflection and the linking of relevant dialogue segments (Figure 1 [B–D]).

Figure 1

Screen Images of EnactingBoard



Conclusion

This study proposed a learning method for sustaining dialogue with emotional opponents and implemented it in the form of the EnactingBoard web application. A future task is to examine the effectiveness of this method through practical use of the system while simultaneously refining the system itself.

 © ISLS

References

Aron, A., Aron, E. N., & Smollan, D. (1992). Inclusion of Other in the Self Scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology*, *63*(4), 596.

Gerace, A., Day, A., Casey, S., & Mohr, P. (2015). Perspective taking and empathy: Does having similar past experience to another person make it easier to take their perspective? *Journal of Relationships Research*, *6*, e10. https://doi.org/10.1017/jrr.2015.6

Hoffman, M. L. (2000). *Empathy and moral development: Implications for caring and justice*. Cambridge University Press.

List, D. (2006). Reflection on the future: Its possibility and usefulness. *Journal of Information Technology Theory and Application*, *7*(4), 5–16.

Morris, M. W., & Moore, P. C. (2000). The lessons we (don't) learn: Counterfactual thinking and organizational accountability after a close call. *Administrative Science Quarterly*, *45*(4), 737–765.

Nickerson, R. S. (1999). How we know—and sometimes misjudge—what others know: Imputing one's own knowledge to others. *Psychological Bulletin*, *125*(6), 737–759.

UNESCO. (2020). *Education for sustainable development: A roadmap*. UNESCO Publishing.

Winslade, J., Monk, G., & Cotte, A. (1998). A narrative approach to the practice of mediation. *Negotiation Journal*, *14*(1), 21–41.

Acknowledgments

This work was supported by JSPS KAKENHI (Grant Numbers 24K00447, 25K21952, 25K00859, and 23K25708).

International Society of the Learning Sciences

Fostering Metacognitive Awareness for Effective Human–AI Collaboration in Learning

Pau Benazet I Montobbio, Universitat Pompeu Fabra, pau.benazeti01@estudiant.upf.edu
Davinia Hernández-Leo, Universitat Pompeu Fabra, davinia.hernandez-leo@upf.edu

Abstract: This exploratory study investigated whether experiential learning is more effective than direct instruction for developing students' metacognitive awareness of AI's effects on learning. Among 126 first-year engineering students, the experimental group completed hands-on exercises demonstrating productive and counterproductive human–AI collaboration, while controls received equivalent lectures. Using bespoke exercises and a novel questionnaire and construct-level analysis, the experiential group showed significant increases in metacognitive awareness and AI engagement, suggesting that direct experience with AI's cognitive impacts is essential for developing reflective human–AI partnerships.

Introduction

The emergence of generative AI (GenAI) based on Large Language Models (LLMs) has fundamentally transformed educational contexts, creating new paradigms for human-AI collaboration in learning (Giannakos et al., 2025; Hernández-Leo et al., 2025). Students are already using these tools regardless of institutional policies (Yan et al., 2024), shifting the educational question from whether to allow AI toward how to help students collaborate with it effectively. Students lack the metacognitive awareness necessary to distinguish between productive and counterproductive AI support (Fan et al., 2025).

Evidence reveals starkly different outcomes depending on human-AI collaboration patterns: inappropriate use can impair memory formation, reduce problem-solving skills, and create an "illusion of competence" (Gerlich, 2025; Kosmyna et al., 2025), while strategic collaboration aligned with learning principles enhances outcomes (Orynbassarova & De Lasala Porta, 2024; Zhang et al., 2024). The ISAR model proposed by Bauer et.al (2025) captures this spectrum, categorizing AI's educational effects as inversion, substitution, augmentation, or redefinition depending on how the tool is integrated into learning.

The challenge is fundamentally metacognitive: students need awareness of how AI interactions affect their cognitive processes, memory encoding, and learning outcomes. Emerging research suggests many students lack this awareness, exhibiting what has been described as metacognitive laziness when using AI tools (Fan et al., 2025). Yet little research has explored how to best develop such awareness. Purely instructional approaches risk triggering psychological reactance (Brehm, 1966) or fostering overconfidence without functional understanding (Castel et al., 2007; Dunning, 2011). Experiential learning theory suggests that concrete experience followed by reflection enables deeper understanding than instruction alone (Kolb et al., 2001).

This exploratory study examines whether experiential learning approaches are more effective than purely instructional approaches for promoting metacognitive awareness on the effects of human-AI collaboration in learning:

- RQ: Is experiential learning on productive and unproductive modes of human-AI collaboration more effective than instructional approaches for promoting metacognitive awareness of how AI partnership affects learning?

Methods

The study followed a pre-post test design conducted in two-hour sessions. Both groups completed the MAI-AI questionnaire, an original instrument based on the Metacognitive Awareness Inventory (Schraw & Dennison, 1994) and the Motivated Strategies for Learning Questionnaire (Duncan & McKeachie, 2005), before and after intervention. The questionnaire consists of 28 items on a 5-point Likert scale across six constructs: Attitudes towards AI, AI Engagement, AI Literacy, Metacognitive Awareness, Metacognitive Regulation, and Autonomy & Control. The full instrument is available at Zenodo (Benazet i Montobbio, 2026).

Participants were (N=126) first-year engineering students from a higher education institution in Spain (after data cleaning from initial N=138), assigned to groups by class schedule: G1 (control, n=61) and G2 (experimental, n=65). The control group (G1) received a traditional lecture explaining productive and unproductive AI collaboration patterns. The experimental group (G2) received the same lecture content but additionally engaged in four structured exercises allowing them to directly experience these patterns. Inspired by the ISAR model, acting as either as an inversion, substitution, augmentation or redefinition effect (Bauer et al.,

 © ISLS

2025), the exercises followed a three-phase structure: (1) perform a task with AI (5–10 min), (2) distraction task (~5 min), (3) similar transfer task without AI (~10 min). A table with a complete description of the exercises can be found at Zenodo (Benazet i Montobbio, 2026).

Results

Between-group differences in composite construct change scores were tested using Mann-Whitney U tests with Bonferroni and Benjamini-Hochberg corrections across six constructs. Two constructs showed significantly different change scores between conditions. Table 1 presents the full results.

Table 1

Between-Group Differences in Composite Construct Change Scores (Post-Pre)

Construct	G1 Mdn (M ± SD)	G2 Mdn (M ± SD)	p_BH	r
Attitudes	0.00 (+0.27 ± 0.46)	0.00 (+0.11 ± 0.76)	.496	-0.101
AI Engagement*	+0.14 (+0.17 ± 0.38)	+0.45 (+0.42 ± 0.41)	**.001**	0.383
AI Literacy	+0.20 (+0.22 ± 0.46)	+0.20 (+0.13 ± 0.50)	.561	−0.074
Metacognitive Awareness*	0.00 (+0.05 ± 0.47)	+0.40 (+0.38 ± 0.61)	**.004**	0.328
Metacognitive Regulation	+0.08 (+0.10 ± 0.41)	+0.17 (+0.09 ± 0.46)	.910	−0.012
Autonomy & Control	0.00 (+0.13 ± 0.57)	0.00 (+0.02 ± 0.54)	.496	−0.098

** Significant after both Bonferroni (p_Bonf = .001, .008) & BH correction. G1: n=61, G2: n=65. Effect size: rank-biserial r.*

The experiential group (G2) showed significantly greater increases in both AI Engagement (r = 0.383) and Metacognitive Awareness (r = 0.328). Notably, the control group's median change in Metacognitive Awareness was zero, indicating that lecture-based instruction alone did not shift awareness of AI's effects on learning. No other constructs differed significantly between groups, including Metacognitive Regulation (p = .910), suggesting the intervention's effects were specific to awareness and engagement rather than a general shift across all dimensions.

Discussion

The construct-level findings provide evidence that experiential engagement with both productive and counterproductive AI collaboration patterns promotes metacognitive awareness more effectively than lecture-based instruction. Students who directly experienced the consequences of different collaboration modes developed stronger awareness of AI's cognitive effects and greater intentions to engage with AI in their learning. These findings connect to work on cognitive offloading (Risko & Gilbert, 2016) and productive failure (Kapur, 2008), suggesting that firsthand experiences is what drives metacognitive development.

The specificity of the effect is worth noting: the intervention increased awareness and engagement but not regulation, attitudes, literacy, or autonomy. This pattern is consistent with metacognition theory, which distinguishes knowledge of cognition from regulation of cognition and recognizes that the latter typically requires more sustained deliberate practice (Flavell, 1979; Pintrich, 2002). Whether and how increased awareness eventually translates into strategic regulation of AI use remains an open question for future research.

Limitations and future work

This exploratory study used a short intervention window (two hours), relied on self-report measures, and the MAI-AI instrument awaits psychometric validation. Group assignment by class schedule limits causal inference. Future research should incorporate extended interventions, behavioral trace data, and psychometric validation of the MAI-AI across larger and more diverse samples.

 © ISLS

References

Bauer, E., Greiff, S., Graesser, A. C., Scheiter, K., & Sailer, M. (2025). Looking Beyond the Hype: Understanding the Effects of AI on Learning. *Educational Psychology Review*, *37*(2), 45. https://doi.org/10.1007/s10648-025-10020-8

Benazet i Montobbio, P. (2026). (Un)Productive Human-AI Collaboration Exercises + MAI-AI Questionnaire - Version 1. Zenodo. https://doi.org/10.5281/zenodo.19243224

Brehm, J. W. (1966). *A theory of psychological reactance* (pp. x, 135). Academic Press.

Castel, A. D., McCabe, D. P., & Roediger, H. L. (2007). Illusions of competence and overestimation of associative memory for identical items: Evidence from judgments of learning. *Psychonomic Bulletin & Review*, *14*(1), 107–111. https://doi.org/10.3758/BF03194036

Duncan, T. G., & McKeachie, W. J. (2005). The Making of the Motivated Strategies for Learning Questionnaire. *Educational Psychologist*, *40*(2), 117–128. https://doi.org/10.1207/s15326985ep4002_6

Dunning, D. (2011). The Dunning–Kruger Effect. In *Advances in Experimental Social Psychology* (Vol. 44, pp. 247–296). Elsevier. https://doi.org/10.1016/B978-0-12-385522-0.00005-6

Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., & Gašević, D. (2025). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology*, *56*(2), 489–530. https://doi.org/10.1111/bjet.13544

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, *34*(10), 906–911. https://doi.org/10.1037/0003-066X.34.10.906

Gerlich, M. (2025). AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking. *Societies*, *15*(1), 6. https://doi.org/10.3390/soc15010006

Giannakos, M., Azevedo, R., Brusilovsky, P., Cukurova, M., Dimitriadis, Y., Hernandez-Leo, D., Järvelä, S., Mavrikis, M., & Rienties, B. (2025). The promise and challenges of generative AI in education. *Behaviour & Information Technology*, *44*(11), 2518–2544. https://doi.org/10.1080/0144929X.2024.2394886

Hernández-Leo, D., Theophilou, E., Oshima, J., Minematsu, T., Matsueda, K., Naganuma, S., Cress, U., Chen, B., Zhu, X., Del Castillo, F. D., Chen, W., Lyu, Q., Chen, X., Zheng, L., & Weinberger, A. (2025). *Collaborating with Generative AI for Learning?* 525–533. https://doi.org/10.22318/cscl2025.882858

Kapur, M. (2008). Productive Failure. *Cognition and Instruction*, *26*(3), 379–424. https://doi.org/10.1080/07370000802212669

Kolb, D. A., Boyatzis, R. E., & Mainemelis, C. (2001). Experiential learning theory: Previous research and new directions. In *Perspectives on thinking, learning, and cognitive styles* (pp. 227–247). Lawrence Erlbaum Associates Publishers.

Kosmyna, N., Hauptmann, E., Yuan, Y. T., Situ, J., Liao, X.-H., Beresnitzky, A. V., Braunstein, I., & Maes, P. (2025). *Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task* (arXiv:2506.08872). arXiv. https://doi.org/10.48550/arXiv.2506.08872

Orynbassarova, D., & De Lasala Porta, S. (2024). Implementing the Socratic Method with AI: Opportunities and Challenges of Integrating ChatGPT into Teaching Pedagogy. *2024 International Conference on Emerging eLearning Technologies and Applications (ICETA)*, 526–532. https://doi.org/10.1109/ICETA63795.2024.10850795

Pintrich, P. R. (2002). The Role of Metacognitive Knowledge in Learning, Teaching, and Assessing. *Theory Into Practice*, *41*(4), 219–225. https://doi.org/10.1207/s15430421tip4104_3

Risko, E. F., & Gilbert, S. J. (2016). Cognitive Offloading. *Trends in Cognitive Sciences*, *20*(9), 676–688. https://doi.org/10.1016/j.tics.2016.07.002

Schraw, G., & Dennison, R. S. (1994). Assessing Metacognitive Awareness. *Contemporary Educational Psychology*, *19*(4), 460–475. https://doi.org/10.1006/ceps.1994.1033

Yan, L., Greiff, S., Teuber, Z., & Gašević, D. (2024). Promises and challenges of generative artificial intelligence for human learning. *Nature Human Behaviour*, *8*(10), 1839–1850. https://doi.org/10.1038/s41562-024-02004-5

Zhang, L., Lin, J., Kuang, Z., Xu, S., & Hu, X. (2024). *SPL: A Socratic Playground for Learning Powered by Large Language Model* (Version 4). arXiv. https://doi.org/10.48550/ARXIV.2406.13919

Acknowledgements

This work has been partly supported by PID2023-146692OB-C33 granted by MICIU/AEI/10.13039/501100011033. D. Hernández-Leo (Serra Húnter) also acknowledges the support of the ICREA Academia programme.

International Society of the Learning Sciences

Hierarchical Multi-Agent System for Instructional Design in Music Knowledge Building

Menglian Zhang, University of York, menglian_zhang@outlook.com
Mingyang Zhu, Nanjing Normal University, zhumy2001@foxmail.com
Haiyang Xin, CocoRobo Ltd, tony@cocorobo.cc
Yibing Zhang, Nanjing Normal University, zhyb304@126.com

Abstract: This study presents a hierarchical multi-agent system (MAS) to support music teachers in designing Knowledge Building (KB)-oriented lessons. The system integrates four expert agents (KB, musician, music pedagogy and music psychology) to generate holistic instructional design that contain embodied and creative music activities. A preliminary qualitative interview with five primary school teachers shows positive feedback of creativity and applicability. Further work will validate the lesson plan quality, compare with single-agent systems, and examine classroom impact through longitudinal studies.

Knowledge Building (KB) in music education

Constructivist in music education emphasises active participation and bodily engagement through authentic, social, and reflective experiences (Shively, 2015). Music education also values culture, affective emotions, and embodied experience (Garnett, 2013; Van Der Schyff et al., 2018). Compared to traditional constructivist education, Knowledge Building (KB) focuses on the continuous and dynamic improvement of ideas through collective learning (Scardamalia & Bereiter, 2010). KB pedagogy encourages deeper learning through artifact production, such as composition, improvisation, and performance in music (Scardamalia & Bereiter, 2010). Collaborative learning, however, has also been applied widely in music teaching and practice (Forbes, 2020), suggesting a potential link between music education and KB pedagogy. Despite these alignments, the direct application of KB pedagogy in music education remains limited. This gap indicates the need for pedagogical and technological innovation to conduct KB principles with creative and affective music learning.

Current issue and exploration in KB music lessons

A primary school in China implementing KB across subjects provide the context for this exploratory study. Five music teachers (4 female, 1 male; 2-20 years of teaching experience; Grades 2, 4, 5 & 6) joined a semi-structured interview addressing theoretical and practical challenges in applying KB and AI tools in music lessons. After thematic analysis, teachers reported difficulty aligning open-ended KB inquiries with curriculum standards (Tan et al., 2021). Some confuse the fixed-group discussion with the collaborative knowledge-sharing groups. Teachers also ask for strategies that connect deeper music understanding and psychological knowledge in pedagogy.

Interdisciplinary skills are needed in KB pedagogy, leading to the use of AI to solve the complex problems in teaching. Existing AI agents often generate too advanced activities for the targeted students' age. For instance, suggested singing tasks exceeded student's vocal abilities, and some embodied activities were impractical with the large students amount in Chinese schools. Teachers further requested clearer inquiry prompts and adaptable activities that support creativity and collaboration. These challenges and requirements highlight the need for greater control through a more flexible AI system that better integrates music practice with KB pedagogy.

Holistic perspective in music teaching and tool use

Teachers prefer an AI system that perform as co-creators, generating synthesised ideas rather than giving concepts and decision. As a complex system, MAS offers such potential by emerge new and unreducible feature through the interaction between individual agents beyond their initial attribute (Chen et al., 2024; Waldrop, 1993). Compared to Large Language Model (LLM) and single agent, MAS can not only reduce produce less hallucination (Huang et al., 2025), but also support complex problems, provide collective intelligence, iterative feedback and flexible subtask management (Li et al., 2024). Accordingly, four experts were designed to reflect the complex and interdisciplinary nature in effective teaching (Mishra & Koehler, 2006), integrating perspectives from music theories, psychology, classroom dynamics, and ethical considerations of AI (Celik, 2023). In detail, KB expert ensures collaborative knowledge creation, while the music educator focuses on music pedagogy and curriculum alignment. The musician contributes to expertise in music theory, history, culture, and performances. Finally, the music psychologist deals with motivational, emotional, and cognitive developmental issues.

The hierarchical Multi-Agent System (MAS)

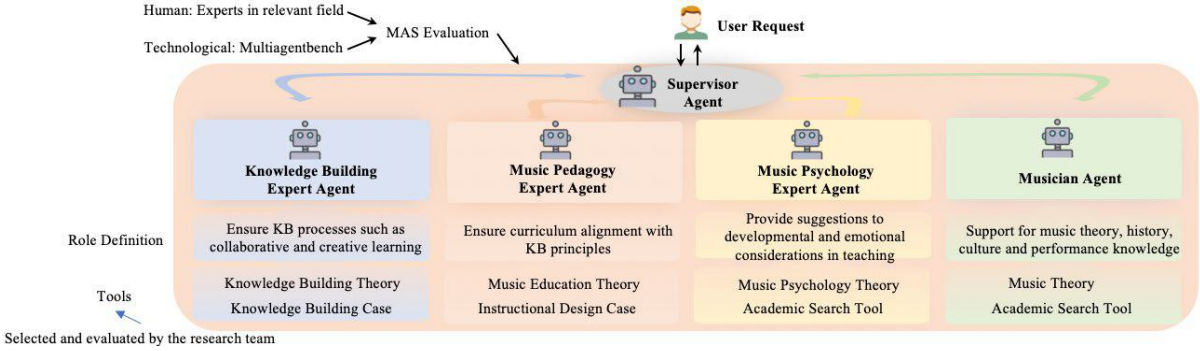
 © ISLS

The hierarchical structure of the Multi-Agent System (MAS) enables modular design and iterative refinement across domains (Talebirad & Nadiri, 2023; Tran et al., 2025). This design aims to generate instructional design (ID) that integrate KB principles with practical music pedagogy, addressing both cognitive and affective aspects, and supports the sophisticated knowledge and teacher control. A supervisor agent manages the overall system goal ($\mathcal{O}_{collab}$), assigns individual goals ($\mathcal{O}_i$) to sub-agents through Standardised Operating Procedures (SOPs). Teachers input teaching objectives or prepared ID through a chat interface. The supervisor agent synthesises a new ID according to the other agents (see Figure 1). Users can then entre requirements to iteratively refine outputs.

The knowledge base contains KB lesson templates, which were validated and adapted by KB experts and music teachers, inquiry prompts, and activity examples supported by musical and psychological literature. Curriculum standards and textbooks ensure practical feasibility.

Figure 1

The Hierarchical MAS for the Generation of Music KB Lesson Plan



Feedback and future plan

System performance was evaluated using Multiagentbench (Zhu et al., 2025). Participants included the 5 music teachers from the previous interview. In a group interview, participants were show about how to input and revise ideas. Three instructional designs about singing and music appreciation for grades 2, 4 and 5 were generated discussed in terms of KB alignment, pedagogical coherence, applicability, developmental appropriateness.

The ID contains KB principles such as collective idea improvement, co-developed artifacts and suggest for student wellbeing. Overall, the new MAS was viewed positively. Teachers found the activities “very creative compared to simple singing activities” and applicable for grade 5 students to move, listen and feel the music (Teacher 2, group interview, September 2025). They also valued the integration of psychological perspectives which support student motivation and engagement. Different from simply a full lesson plan provided by another MAS developed by Xin et al. (2025), this hierarchical MAS can more adequately capture the affective dimensions.

However, teachers requested clearer output structures: “The content and activities were basically feasible, but can you make the output clearer—follow the template we always use?” (Teachers 3, group interview, September 2025). Specifically, a novice teacher asked, “I want more inquiries in the lesson plan, better to cover every sentence I need to say in class” (Teacher 5, group interview, Septermber 2025). Experienced teachers agreed this value, especially for preparing demonstration lessons (Teachers 2 and 4, group interview, September 2025).

The development of this system and verification of teaching effectiveness have not been complete. More lesson plans revised by experts will be added to the knowledge base to improve comprehensiveness. A validated evaluation scale based on the ADDIE model (Branch, 2009) will be created for experts in KB and music psychologist to scoring. The usability, teacher satisfaction, and learning outcomes will be examined through longitudinal studies. Technical improvement will address reaction speed, stream output, and interface design.

Conclusion

This study presents a preliminary design of a complex hierarchical MAS for KB informed music instructional design. By integrating KB theory, music pedagogy and psychology, the system offers a more interactive and flexible approach which supports creative, collaborative, and student-centred learning. While further validation is required, findings suggest that MAS can enhance ID while teachers are encouraged to adapt AI-generating lesson plans to actual classroom and cultural contexts to maintain subjectivity (Hesse & Helm, 2025; Pikhart & Al-Obaydi, 2025). This MAS design can be further expanded to interdisciplinary learning and practical music skills, including vocal and instrumental performances, particularly in Chinese educational context.

Reference

Branch, R. M. (2009). *Instructional Design: The ADDIE Approach*. Springer US. https://doi.org/10.1007/978-0-387-09506-6

Celik, I. (2023). Towards Intelligent-TPACK: An empirical study on teachers' professional knowledge to ethically integrate artificial intelligence (AI)-based tools into education. *Computers in Human Behavior*, *138*, 107468. https://doi.org/10.1016/j.chb.2022.107468

Chen, W., Su, Y., Zuo, J., Yang, C., Yuan, C., Chan, C.-M., Yu, H., Lu, Y., Hung, Y.-H., Qian, C., Qin, Y., Cong, X., Xie, R., Liu, Z., Sun, M., & Zhou, J. (2024, May). *AGENTVERSE: FACILITATING MULTI-AGENT COLLAB- ORATION AND EXPLORING EMERGENT BEHAVIORS*. In The Twelfth International Conference on Learning Representations, Austria.

Li, X., Wang, S., Zeng, S., Wu, Y., & Yang, Y. (2024). A survey on LLM-based multi-agent systems: Workflow, infrastructure, and challenges. *Vicinagearth*, *1*(1), 9. https://doi.org/10.1007/s44336-024-00009-2

Forbes, M. (2020). The value of collaborative learning for music practice in higher education. *British Journal of Music Education*, *37*(3), 207–220. https://doi.org/10.1017/S0265051720000200

Garnett, J. (2013). Beyond a constructivist curriculum: A critique of competing paradigms in music education. *British Journal of Music Education*, *30*(2), 161–175. https://doi.org/10.1017/S0265051712000575

Hesse, F., & Helm, G. (2025). Writing with AI in and beyond teacher education: Exploring subjective training needs of student teachers across five subjects. *Journal of Digital Learning in Teacher Education*, *41*(1), 21–36. https://doi.org/10.1080/21532974.2024.2431747

Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., Chen, Q., Peng, W., Feng, X., Qin, B., & Liu, T. (2025). A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions. *ACM Transactions on Information Systems*, *43*(2), 1–55. https://doi.org/10.1145/3703155

Mishra, P., & Koehler, M. J. (2006). Technological Pedagogical Content Knowledge: A Framework for Teacher Knowledge. *Teachers College Record: The Voice of Scholarship in Education*, *108*(6), 1017–1054. https://doi.org/10.1111/j.1467-9620.2006.00684.x

Pikhart, M., & Al-Obaydi, L. H. (2025). Reporting the potential risk of using AI in higher Education: Subjective perspectives of educators. *Computers in Human Behavior Reports*, *18*, 100693. https://doi.org/10.1016/j.chbr.2025.100693

Scardamalia, M., & Bereiter, C. (2010). A Brief History of Knowledge Building. *Canadian Journal of Learning and Technology / La Revue Canadienne de l'apprentissage et de La Technologie*, *36*(1). https://doi.org/10.21432/T2859M

Shively, J. (2015). Constructivism in Music Education. *Arts Education Policy Review*, *116*(3), 128–136. https://doi.org/10.1080/10632913.2015.1011815

Talebirad, Y., & Nadiri, A. (2023). *Multi-Agent Collaboration: Harnessing the Power of Intelligent LLM Agents* (No. arXiv:2306.03314). arXiv. https://doi.org/10.48550/arXiv.2306.03314

Tan, S. C., Chan, C., Bielaczyc, K., Ma, L., Scardamalia, M., & Bereiter, C. (2021). Knowledge building: Aligning education with needs for knowledge creation in the digital age. *Educational Technology Research and Development*, *69*(4), 2243–2266. https://doi.org/10.1007/s11423-020-09914-x

Tran, K.-T., Dao, D., Nguyen, M.-D., Pham, Q.-V., O'Sullivan, B., & Nguyen, H. D. (2025). *Multi-Agent Collaboration Mechanisms: A Survey of LLMs* (No. arXiv:2501.06322). arXiv. https://doi.org/10.48550/arXiv.2501.06322

Van Der Schyff, D., Schiavio, A., Walton, A., Velardo, V., & Chemero, A. (2018). Musical creativity and the embodied mind: Exploring the possibilities of 4E cognition and dynamical systems theory. *Music & Science*, *1*, 2059204318792319. https://doi.org/10.1177/2059204318792319

Waldrop, M. M., & Waldrop, M. M. (1993). *Complexity: The emerging science at the edge of order and chaos* (1. Touchstone ed). Touchstone.

Xin, H., Lan, L., Niu, Q., Hu, Z., & Zhang, Y. (2025). *A Multi-Agent System (MAS)-Based Tool to Support Novice Teachers in Knowledge Building Pedagogy*. 617–618. https://doi.org/10.22318/cscl2025.605711

Zhu, K., Du, H., Hong, Z., Yang, X., Guo, S., Wang, Z., Wang, Z., Qian, C., Tang, X., Ji, H., & You, J. (2025, July-August). *MultiAgentBench: Evaluating the Collaboration and Competition of LLM agents*. In the 63rd Annual Meeting of the Association for Computational Linguistics,Austria.

International Society of the Learning Sciences

AI Tools in Arts Learning: A Scoping Review

Jaycee Sansom, Safinah Ali
js15197@nyu.edu, safinah.ali@nyu.edu
New York University

Abstract: With the proliferation of Artificial Intelligence (AI) tools in art making, AI knowledge and skills have made their way to arts learning. However, an assessment of the current state of AI tools in arts learning is lacking. This study used scoping review methodology to identify current research trends in AI applications in arts learning, and what gaps exist for future research. 58 papers were reviewed revealing four trends and five gaps and opportunities.

Introduction

With advancements in generative AI, AI tools and techniques have especially found applications in art making - and have subsequently been introduced in art learning (Ali, 2025). Gaining AI skills and knowledge becomes increasingly relevant to art learners - both as AI finds applications in art making, and as AI tools impact artists' careers (Jiang, et al., 2023). However, as found by Zhai, et al. (2021), between 2010 and 2020, few papers were focused on AI in *arts* education and learning. This is concerning because artists can benefit significantly from AI literacy initiatives by expanding their skillsets, confidence, and even sense of belonging in the field, as well as better preparing themselves for a future where AI technologies can adversely impact creative careers.

As such, it is important to recognize the trends and opportunities in AI and arts learning that exist so we may better understand where support is needed, as well as to identify where gaps in practice exist. More specifically, we wish to answer the following research questions: *what areas of arts learning are being focused on, how is AI being used and looked at in these studies, what are the common themes present in the current state of research, and what gaps and opportunities exist for future research?* To address this need, a scoping review is needed to examine the current state of research in AI tools in arts learning. We define arts learning as "a field consisting of creative studies including visual arts, music, creative writing, theater, dance, fabric arts, and other disciplines". Following this definition and the scoping review methodology we identified 58 articles for review, resulting in four current trends in arts education and AI research, and five gaps in knowledge and applications.

Methodology

In order to systematically evaluate the state of AI tools in arts learning, a scoping review was conducted, following a structured process. A total of 548 papers were screened from AI and education journals and conference proceedings. We used - Google Scholar, ISLS's database, and the NYU library database to conduct the search. Papers that met the following exclusion criteria were then excluded from analysis –: published before 2019 (due recency and relevance), not focused on arts learning or AI, a literature review, not peer reviewed, workshop papers, perspectives papers (no original research), duplicate record, or were not indexed or published in known journals or databases (e.g. Scopus, ERIC, DOAJ, etc.). The 58 papers that met the criteria were analyzed using multiple rounds of coding based on grounded theory (Corbin & Strauss, 2014). Codes were inductively generated through a two-round coding process where themes were identified and refined to answer our research questions. One category of code, methodology, was based on codes developed by Zhai, et al., 2021.

Results

Six categories of non-exclusive codes were identified through the thematic analysis process that informed the research questions of this work: "Area of Arts Learning," "AI Applications," "Education Level," "Study Objectives," "Learner Objectives," and "Methodology." Several studies had multiple codes in one category.

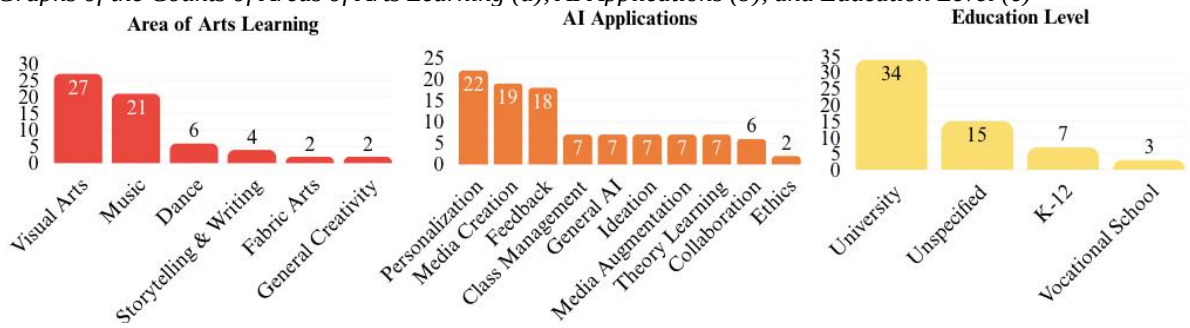
The most commonly explored areas of art learning in the reviewed studies were visual arts (47% of papers) and music (36%) (Figure 1.a). Fewer studies (22%) focused on dance, storytelling and writing, fabric arts, and general creativity, and no studies were identified in the theater domain, even with specific term elicitation. Common AI applications that were focused upon included adaptive systems for personalization to learners (38%), media creation (33%), and feedback (31%) (Figure 1.b). The next group of applications – class management, ideation, theory learning, collaboration, and media augmentation were utilized as well in some studies, but not nearly as often as the three most common applications. The next category of codes that emerged was the education level of the participants (Figure 1.c). A vast majority (59%) of the studies examined focused on university students and professors, with K-12 and vocational schools having much smaller representation (12% and 5%). Three categories for study objectives were identified: perspective studies, experiment studies, and system proposals.

 © ISLS

Learner objectives evaluated in the reviewed studies most commonly included domain-specific skills (34%), which was where studies assessed student acquisition of specific skills, such as, design thinking or technical art skills, as opposed to general performance and achievement (19%). Other common themes for this category included: motivation, performance, and learner perspectives about AI. Finally, there were some trends found with the analysis methodology used by the studies. Experimental and quasi-experimental methods were used the most frequently (64%), followed by surveys (48%).

Figure 1

Graphs of the Counts of Areas of Arts Learning (a), AI Applications (b), and Education Level (c)



(a) (b) (c)

Discussion and conclusion

Based on the findings, a total of four current trends in research were identified. Current research in AI in art learning primarily focuses on personalized learning and automated feedback through the use of AI. Personalized learning tended to involve creating custom content and resources (e.g. Chen & Qu, 2025) or the personalization of learning pathways and resource recommendations typically through the use of big data and analytics (e.g. Deng & Wang, 2024). Automated feedback provided real-time or quick feedback to students in place of teachers (e.g. Chen, 2020). Another trend we observed is the use of more formal assessment tools to evaluate students after having used AI technology. This often involved using a pre-established method for assessing creativity and overall performance, such as the Torrance Test of Creative Thinking (e.g. Wang, 2025) or relying on experts and teachers (e.g. Kara, 2025). Current research also revealed that art students found AI tools more useful as an assistant or a collaborator in the early stages of the creative process, rather than during the creation of the final artifact (e.g. Ansone, et al., 2025). Studies such as these may imply that art students are aware of the limitations of generative AI as a creator of an end product, but it can serve as a useful tool for exploring creative interpretations of a task. Finally, perspective studies of students and teachers were a notable trend. This often took the form of one of two types of studies. The first type were studies that look exclusively at the perspectives of students and teachers (e.g. Bedir Erişti & Freedman, 2024) while the second type consisted of studies that introduced an AI tool and looked at students' perspectives regarding the tool (e.g. Deval, et al., 2025).

Five notable gaps in the literature were identified. Studies examined in this review reported having a limited research scope, including small population sizes (e.g. Keune, et al., 2024), lack of a long-term evaluation (e.g. Natheer, 2024), or lack of diversity in either the population (e.g. Zhang, et al., 2025) or the technology used (e.g. Zhang, 2025). Similarly, a vast majority of the papers examined focused on feedback, and personalization in art learning, while applications such as collaboration and ethics remain less explored. Furthermore, there is a scope of exploring more AI technologies that can benefit art learning beyond large language models. Additionally, there were papers that proposed a novel approach to an algorithm or other AI implementation that can benefit art learning (e.g. Jiang, et al., 2025). Few of these studies went beyond benchmarking their algorithms/systems and explored user studies, which in itself presents an opportunity for future research. There also appears to be a noticeable gap in teacher involvement. Teachers in the examined studies, if involved, were only involved in one of two ways, as perspectives (e.g Bedir Erişti & Freedman, 2024) or as assessors of student performance (e.g. Chen, et al., 2025). The final gap present in this paper is that few works focus on *how* students are learning from AI tools in the arts. Learning outcomes were commonly examined through design interventions, but oftentimes, the learning or art creation process itself was not investigated.

By exploring gaps and opportunities in the area, this work highlights opportunities for research in the fields of AI literacy, education technology, and the learning sciences. This review is not without its limitations as the studies selected were influenced by the search query used and inter-coder reliability was not calculated.

 © ISLS

References

Ali, S. (2025). Designing AI Tools to Support Art Learning: A Case Study of Four AI Tools to Support Creative Inquiry. In *Constructionism Conference Proceedings* (Vol. 8, pp. 341-350).

Ansone, A., Zālīte-Supe, Z., & Daniela, L. (2025). Generative Artificial Intelligence as a Catalyst for Change in Higher Education Art Study Programs. *Computers (Basel)*, 14(4), Article 154. https://doi.org/10.3390/computers14040154

Bedir Erişti, S. D., & Freedman, K. (2024). Integrating Digital Technologies and AI in Art Education: Pedagogical Competencies and the Evolution of Digital Visual Culture. *Participatory Educational Research*, 11(H. Ferhan Odabaşı Gift Issue), 57–79. https://doi.org/10.17275/per.24.94.11.6

Chen, J. C. W. (2020). AI in music education: The impact of using artificial intelligence (AI) application to practise scales and arpeggios in a virtual learning environment. In *Learning Environment and Design: Current and Future Impacts* (pp. 307-322). Singapore: Springer Singapore.

Chen, J., Mokmin, N. A. M., & Su, H. (2025). Integrating generative artificial intelligence into design and art course: Effects on student achievement, motivation, and self-efficacy. *Innovations in Education and Teaching International*, 1431-1446.

Cheng, D., & Qu, X. (2025). Intelligent generation and optimization of resources in music teaching reform based on artificial intelligence and deep learning. *Scientific Reports*, 15(1), Article 30051. https://doi.org/10.1038/s41598-025-16458-8

Corbin, J., & Strauss, A. (2014). Basics of qualitative research: Techniques and procedures for developing grounded theory. *Sage publications*.

Deng, K., & Wang, G. (2024). Online mode development of Korean art learning in the post-epidemic era based on artificial intelligence and deep learning. *The Journal of Supercomputing*, 80(6), 8505-8528. https://doi.org/10.1007/s11227-023-05776-1

Deval, Ö., Öztürk Kösenciğ, K., & Acırlı, Z. (2025). Integration of artificial intelligence tools into interior architecture education: a study on textual and visual representations. *Estoa*, 14(28), 109–120. https://doi.org/10.18537/est.v014.n028.a08

Jiang, H. H., Brown, L., Cheng, J., Khan, M., Gupta, A., Workman, D., Hanna, A., Flowers, J., & Gebru, T. (2023). AI Art and its Impact on Artists. In *AIES 2023 - Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 363–374). ACM. https://doi.org/10.1145/3600211.360468

Jiang, S., Shi, N., & Liu, C. (2025). The analysis of artificial intelligence knowledge graphs for online music learning platform under deep learning. *Scientific Reports*, 15(1), Article 16481. https://doi.org/10.1038/s41598-025-01810-9

Kara, S. (2025). The effect of artificial intelligence applications in 6th grade visual arts course on student attitudes and course outcomes. *International Journal of Modern Education Studies*, 9(1). https://doi.org/10.51383/ijonmes.2024.405

Keune, A., Hurtado, S., & Simšič, Ž. (2024). Exploring AI Ethics through educational scenarios with AI generative arts apps. In *Proceedings of the 18th International Conference of the Learning Sciences-ICLS 2024*, pp. 1818-1821. International Society of the Learning Sciences.

Natheer, W. I. (2024). The Impact of Dream Artificial Intelligence Relevance and Design Skills Proficiency on Art Education Students in Designing Innovative Decorative Units. *Educational Sciences: Theory & Practice*, 24(2), 107-121. https://doi.org/10.12738/jestp.2024.2.008

Wang, X. (2025). Integration of AI GPTs in music education and their impact on students' perception and creativity. *Education and Information Technologies*, 30(13), 18879–18893. https://doi.org/10.1007/s10639-025-13531-7

Zhai, X., Chu, X., Chai, C. S., Jong, M. S. Y., Istenic, A., Spector, M., Liu, J.-B., Yuan, J., Li, Y., & Cai, N. (2021). A Review of Artificial Intelligence (AI) in Education from 2010 to 2020. *Complexity*, 2021(1), Article 8812542. https://doi.org/10.1155/2021/8812542

Zhang, L. (2025). Compositional tools based on artificial intelligence for choral artistic education: Enhancing creative skills in choral arrangements. *Thinking Skills and Creativity*, 56, Article 101768. https://doi.org/10.1016/j.tsc.2025.101768

Zhang, R., Lee, H., Wu, R., Yang, W., & Pan, Y. (2025). Is the impact of artificial intelligence generation tools on improving art education positive or negative? Perspectives of professors and students. *Interactive Learning Environments*, 33(10), 5944–5975. https://doi.org/10.1080/10494820.2025.2490173

International Society of the Learning Sciences

Principle-Driven Scaffolding: Implementing and Evaluating LLM-Generated Micro-Scripts in an Informal Learning Community

Hideki Kondo, Kanda University of International Studies, kondo-h@kanda.kuis.ac.jp

Abstract: How can AI scaffold collaboration in voluntary online communities without over-scripting participation? This poster presents Principle-Driven Micro-Script Generation (PDMSG), an LLM-based approach that translates activity objectives, local constraints, learning principles, and recent community activity logs into optional weekly suggestions for each member. In a seven-week deployment of PDMSG in BookClub, text volume per Presence increased without a corresponding rise in post-count density, and participants rated the suggestions as readable and minimally coercive.

Introduction

Informal online communities need support that encourages collaborative inquiry without turning participation into compliance. In CSCL, collaboration scripts have long been a major approach to providing such guidance by structuring roles, sequences, and procedures for interaction (Kollar et al., 2006; Fischer et al., 2013). At the same time, too much scripting may undermine agency, especially in voluntary and asynchronous settings where goals, topics, and participation shift over time (Dillenbourg, 2002; Johnson & Majewska, 2022). Recent work on generative AI, shared control, and AI-supported regulation suggests that this issue remains salient as AI enters collaborative learning design (Cress & Kimmerle, 2023; Hernández-Leo et al., 2025; Kim et al., 2025).

To implement this idea, a framework called Principle-Driven Micro-Script Generation (PDMSG) has been proposed (Kondo, 2025). PDMSG integrates four elements into a single prompt for an LLM to generate context-aware micro-scripts for each member: (1) activity objectives, (2) local constraints, (3) learning principles, and (4) recent activity logs.

The initial PDMSG proposal tested a script-generation method grounded in the 12 Knowledge Building principles. It showed that the generated scripts could be context-relevant, actionable, and theory-aligned, but it did not evaluate whether the method functioned in a live setting. This poster reports a field deployment in BookClub and focuses on the feasibility, acceptability, and behavioral tendencies associated with its introduction.

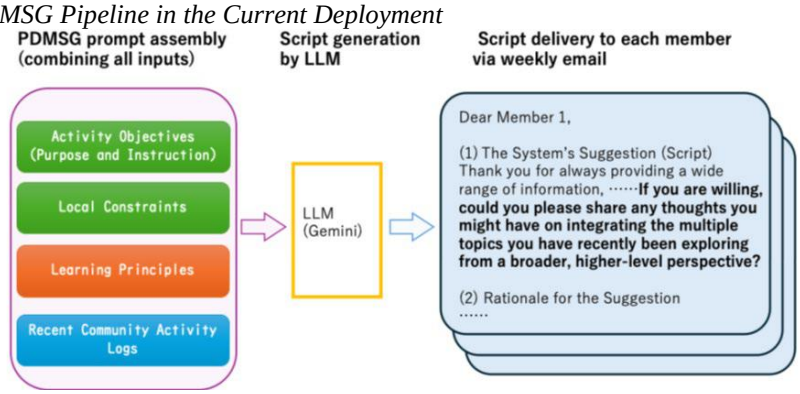
Principle-Driven Micro-Script Generation

Figure 1 presents the PDMSG pipeline. The shared inputs were the activity objectives, local constraints, learning principles, and recent community activity logs. These inputs were combined to generate, for each member, the collaboration script itself and an explanation of the rationale for its generation.

In this deployment, the learning-principles slot was instantiated with the 12 Knowledge Building (KB) principles. The unit of generation was the individual member. Because BookClub is a voluntary and emergent community, the prompt did not assign stable roles such as questioner or summarizer, as fixed role allocation risked over-scripting participation. Still, the generated suggestions could emphasize moves analogous to questioning, synthesizing, or linking ideas, depending on recent logs and the shared community context.

Figure 1

PDMSG Pipeline in the Current Deployment



 © ISLS

Method

We employed a single-group pre-post comparison design, establishing a seven-week intervention period within the active BookClub. BookClub is a small-scale, long-running informal reading community that uses a threaded discussion board and holds weekly synchronous Zoom meetings. BookClub is oriented toward the sustained improvement of ideas rather than predefined performance goals, making it a useful setting for KB-grounded scaffolding (Kondo, 2025; Scardamalia & Bereiter, 1994; Scardamalia, 2002). Six members (excluding the facilitator) participated.

During the intervention, every Monday morning, we generated a unique collaboration script and its rationale for each member and sent it by email, explicitly stating that members were free to ignore the suggestion. The quantitative log analysis focused on the asynchronous discussion board. We defined Presence as the weekly sum of daily visits to the discussion board, capped at two per day, and analyzed posts and characters per unit of Presence ("presence density") across the baseline and intervention periods. Following the intervention, we administered a survey with 5-point Likert items and open-ended questions; open-ended responses were reviewed descriptively. This study does not claim causal effects; it examines feasibility, acceptability, and associated behavioral tendencies. The Likert items were designed to probe script acceptability, feasibility under busyness, and perceived changes in the KB-consistent activities that the intervention was intended to support.

Results

The main pattern in the log data was not in post-count density, but in the volume of content per unit of Presence. New-text density increased from 39.8 to 44.6 characters per Presence (RR = 1.119), and reply-text density increased from 71.1 to 100.8 (RR = 1.418), while post-count density remained stable or slightly lower. At the individual level, the text volume of new posts increased for all six participants, and the text volume of replies increased for five of the six members.

Survey responses were broadly consistent with this pattern. Readability was high (4.7/5), while specificity (3.8/5) and feasibility under busyness (3.7/5) were moderate to high; perceived coerciveness was low (2.0/5). Members also reported increases, particularly in following up on existing threads and in summarizing or organizing ideas. Contextual fit was rated moderately (3.0/5), and one open-ended response noted that "[The system] treated me like an expert in an area I had only mentioned briefly, which made me feel pressured and unable to write anything."

Discussion

This study examined whether PDMSG was acceptable to members and associated with KB-oriented participation patterns in BookClub. The clearest shift was not in post-count density but in text volume per unit of Presence.

Taken together with participants' self-reports, this pattern may reflect more elaborated contribution practices, especially following up on existing threads and summarizing or organizing ideas. The low perceived coerciveness ratings provide no clear indication of perceived coercion in this setting. However, this should not be taken as a full measure of epistemic agency.

At the same time, this small-scale, seven-week study of six participants lacked a comparison condition, so the results indicate association rather than causation; specifically, we cannot separate PDMSG effects from those of regular weekly suggestion emails. While text volume per Presence may be a useful proxy here, finer-grained analysis remains future work. Open-ended responses suggest two design improvements: weighting recent logs more heavily and linking to the posts underlying each script. Embedding this support in a KB-oriented environment such as Knowledge Forum may make each suggestion's basis more visible, though whether that would change the pattern remains an open question.

Conclusion

This seven-week study provides initial indications of the feasibility and acceptability of the PDMSG framework in this setting. In an implementation that incorporated the 12 Knowledge Building principles, system logs revealed a consistent increase in text volume per Presence, and survey responses correspondingly suggested increases in activities such as following up on, summarizing, and organizing existing ideas. Taken together, these findings suggest that PDMSG may help link collaboration scripts with KB-oriented activity and support the translation of theory into situated practice.

For the poster session, we invite discussion on (1) how much pipeline transparency is needed to make PDMSG reusable; (2) what more direct and fine-grained measures of epistemic depth and agency should complement presence density and perceived coerciveness; and (3) how this kind of support might be embedded in KB-oriented platforms.

References

Cress, U., & Kimmerle, J. (2023). Co-constructing knowledge with generative AI tools: Reflections from a CSCL perspective. International Journal of Computer-Supported Collaborative Learning, 18(4), 607–614. https://doi.org/10.1007/s11412-023-09409-w

Dillenbourg, P. (2002). Over-scripting CSCL: The risks of blending collaborative learning with instructional design. In P. A. Kirschner (Ed.), Three worlds of CSCL: Can we support CSCL? (pp. 61–91). Open University of the Netherlands.

Fischer, F., Kollar, I., Stegmann, K., & Wecker, C. (2013). Toward a script theory of guidance in computer-supported collaborative learning. Educational Psychologist, 48(1), 56–66. https://doi.org/10.1080/00461520.2012.748005

Hernández-Leo, D., Theophilou, E., Oshima, J., Minematsu, T., Matsueda, K., Naganuma, S., Cress, U., Chen, B., Zhu, X., del Castillo, F. D., Chen, W., Lyu, Q., Chen, X., Zheng, L., & Weinberger, A. (2025). Collaborating with Generative AI for learning? In J. Oshima, B. Chen, F. Vogel, & J. Järvelä (Eds.), Proceedings of the 18th International Conference on Computer-Supported Collaborative Learning—CSCL 2025 (pp. 525–533). International Society of the Learning Sciences. https://doi.org/10.22318/cscl2025.882858

Johnson, M., & Majewska, D. (2022). Formal, non-formal, and informal learning: What are they, and how can we research them? Cambridge University Press & Assessment. https://doi.org/10.17863/CAM.110841

Kim, J., Detrick, R., Yu, S., Song, Y., Bol, L., & Li, N. (2025). Socially shared regulation of learning and artificial intelligence: Opportunities to support socially shared regulation. Education and Information Technologies, 30, 11483–11521. https://doi.org/10.1007/s10639-024-13187-9

Kollar, I., Fischer, F., & Hesse, F. W. (2006). Collaboration scripts—A conceptual analysis. Educational Psychology Review, 18(2), 159–185. https://doi.org/10.1007/s10648-006-9007-2

Kondo, H. (2025). Principle-Driven Micro-Script Generation for informal collaborative learning on discussion boards. IIAI Letters on Informatics and Interdisciplinary Research, 6. https://doi.org/10.52731/liir.v006.398

Scardamalia, M. (2002). Collective cognitive responsibility for the advancement of knowledge. In B. Smith (Ed.), Liberal education in a knowledge society (pp. 67–98). Open Court.

Scardamalia, M., & Bereiter, C. (1994). Computer support for knowledge-building communities. The Journal of the Learning Sciences, 3(3), 265–283. https://doi.org/10.1207/s15327809jls0303_3

Acknowledgments

This work was supported by JSPS KAKENHI Grant-in-Aid for Scientific Research 24K06334.

International Society of the Learning Sciences

Equitable Learning Analytics Dashboards: Roles and Mechanisms in Personalized Learning

Wei Yan, University of Texas at San Antonio, wei.yan@my.utsa.edu
Hyejeong Lee, University of Texas at San Antonio, hyejeong.lee@utsa.edu
Donna Dancer, University of Texas at San Antonio, donna.dancer@my.utsa.edu
Wonjin Yu, University of Alabama, wyu21@ua.edu

Abstract: This poster presents a rapid scoping review of equitable learning analytics dashboards (LADs) for personalized learning. Grounded in the integration of self-regulated learning (SRL) theory and situated learning theory, the review synthesizes findings from ten journal articles published between 2023 and 2025. The findings outline four roles and eight supporting mechanisms through which personalized LADs promote equity. This preliminary synthesis highlights inclusive design consideration to support equitable LAD development.

Introduction and theoretical framework

Learning analytics dashboards (LADs) are sophisticated interfaces that integrate and visualize data about learning processes to support educational decision-making. Personalized learning in the context of LADs involves presenting dynamically generated data to support adaptation of study strategies and pacing, aiming to foster self-regulated learning (SRL) competencies (Verbert et al., 2020). With the development of technology, Artificial intelligence (AI) analytics and multimodal data integration have significantly enhanced LAD capabilities, yet critical gaps remain in synthesizing their equity impacts (Baker et al., 2022; Wise & Shaffer, 2023). To address these research gaps, this review synthesizes recent studies on how personalized LADs promote equity. The research question guided this study is: *What are the roles of personalized LADs in promoting equity, and what technical mechanisms support these roles?*

The theoretical framework integrates self-regulated learning (SRL) theory (Zimmerman, 2002) and situated learning theory (Lave & Wenger, 1991). SRL conceptualizes effective learning as iterative cycles of forethought, performance, and self-reflection, enabling learners to actively regulate and refine their learning strategies (Zimmerman, 2002). In the context of LADs, SRL provides a lens to examine how data visualizations and adaptive feedback empowers students to take control over their learning (Lee, 2024). Situated learning theory posits that knowledge is constructed through participation in authentic social contexts, where learners progressing from legitimate peripheral participation toward full membership in communities of practice (Lave & Wenger, 1991). In the context of LADs, this perspective underscores how LADs foster equity by scaffolding inclusive participation and engagement. By combing SRL theory and situated learning theory, this study adopted a dual lens to explore how personalized LADs promote equity.

Methods

This study adopted a rapid scoping review method to systematically map and synthesize relevant literature. As an accelerated form of the traditional review, this method streamlines processes to deliver timely insights into fast evolving fields (Tricco et al., 2015). It is well-suited to educational technology because rapid technological advances call for up-to-data review that capture emerging trends and developments.

This study followed the four steps: (1) Identification: This study followed the Preferred Reporting Items for Systematic Reviews (PRISMA) framework (Page et al., 2021) to identify related studies. Searches were conducted in three databases: Scopus, IEEE Xplore, and web of science. The Boolean search string used was "*learning analytics dashboards*" AND "*personalized learning*" AND "*equity*" OR "*inclusion*" with filters applied to limit results to peer-reviewed articles. (2) Screening**:** Title and abstracts were independently screened by four researchers against the inclusion criteria: (a) published in English; (b) published between 2023 to 2025; (c) focus on personalized learning in LADs; (d) relate to equity; and (e) published in top-tier Q1 academic journals ranked by the Scimago Journal Rank (SJR). (3) Coding: Data extraction was performed using NVivo software with a deductive-inductive coding framework grounded in SRL theory and situated learning theory. And (4) Analysis: Coding was aggregated through thematic synthesis to identify patterns. Four researchers independently coded and analyzed the data, with discrepancies resolved through discussion and consensus.

Findings

The findings were organized into four roles and eight supporting technical mechanisms that facilitate equitable implication for diverse learners (see Table 1).

 © ISLS

- **Providing data-driven insights**. Personalized LADs provide students with data-driven insights (Yan et al., 2025). They promote equity by offering tailored support and SRL profiles that foster diverse learner agency. For example, Abouelenein et al. (2025) utilized Moodle's learning logs to track students goal setting and questions answered, generating visualizations that enabled automated guidance to adjust learning paths. Similarly, Giorgashvili et al. (2025) leveraged Moodle-integrated LADs to monitor real-time interactions and progress, offering personalized guidance to foster self-reflection in writing tasks.
- **Optimizing feedback delivery**. Personalized LADs provide timely and actionable feedback that enhances students' self-reflection and adaptive learning behaviors. They promote equity by delivering personalized, context-sensitive feedback and bridging accessibility to ensure learners with diverse needs can understand and act on it (Lee et al., 2025). The technological mechanism involves hierarchical content delivery supported by adaptive LADs that monitor progress. For example, Maier and Klotz (2025) used MasteryX adaptive grammar app, which organizes learning into hierarchical course levels and provides elaborated feedback that enable students to receive actionable feedback.
- **Advancing adaptive personalization**. Personalized LADs enable scalable, individualized learning pathways that adapt to diverse learner needs. This role is supported by technological mechanism such as AI-driven personalization and natural language processing (NLP). For example, Tretow-Fish et al. (2023) demonstrated the use of AI for evaluation and adaptative content delivery. Similarly, Weidlich et al. (2025) employed NLP to analyze Moodle forum interaction and identify indicators such as academic terminology usage, facilitating responsive and equitable support.
- **Integrating multimodal data**. Personalized LADs process diverse multimodal inputs such as text, video, audio to address different learner backgrounds and accessibility needs. Key supporting mechanisms include multimodal data processing and bias-aware algorithmic design. For example, Bauer et al. (2025) analyze multimodal learner data, including log files, eye tracking, videos, and texts, to enable responsive and inclusive support. Küchemann et al. (2025) employed large multimodal models as middleware to facilitate real-time scaffolding. Similarly, Wu and Chiu (2025) explored the integration of generative AI tools, including ChatGPT for text generation, Midjourney for image creation, and Grammarly for writing assistance, to deliver interactive feedback directly through user interactions.

Table 1
Roles and Mechanisms of Personalized LADs in Fostering Equity

Roles	Equity Implication	Technological Mechanism	Supported Studies
Providing data-driven insights	• SRL profiles supporting diverse learner agency • Tailored interventions for learner diversity	• Real-time LMS dashboards monitoring • Data-driven algorithms for personalization	Abouelenein et al. (2025) Giorgashvili et al. (2025)
Optimizing feedback delivery	• Personalized and equitable feedback delivery • Bridging accessibility and comprehension gap	• Mastery-based adaptive tracking • Message or GenAI-based feedback delivery	Jung & Wise (2025) Maier & Klotz (2025)
Advancing adaptive personalization	• Scalable personalization for diverse learners • Inclusive data processing	• AI/ML-driven personalization • Natural language processing	Ning et al. (2025) Tretow-Fish et al. (2023) Weidlich et al. (2025)
Integrating multimodal data	• Algorithmic biases mitigation • Equitable learner engagement and interaction	• Large multimodal model • Algorithmic biases mitigation	Bauer et al. (2025) Küchemann et al. (2025) Wu & Chiu (2025)

Significance and future research

This review contributes to the literature by synthesizing recent studies on personalized LADs and their roles in promoting equity. By integrating the perspectives of SRL and situated learning, this study highlights how personalized LADs function not only as tools but also as mechanisms that foster inclusive participation in learning environments. Future research should further examine the long-term impact of personalized LADs on learners across diverse contexts and investigate potential biases embedded in AI-driven analytics systems. In addition, mixed-methods approaches are needed to better understand how students interpret dashboard feedback and use these insights to regulate their learning.

References

The symbol of * represents the studies included in this review.

*Abouelenein, Y. A. M., Selim, S. A. S., & Aldosemani, T. I. (2025). Impact of an adaptive environment based on learning analytics on pre-service science teacher behavior and self-regulation. *Smart Learning Environments*, 12(1).

Baker, R. S., & Hawn, A. (2022). Algorithmic Bias in Education. *International Journal of Artificial Intelligence in Education*, *32*(4), 1052–1092.

* Bauer, E., Heitzmann, N., Bannert, M., Chernikova, O., Fischer, M. R., Frenzel, A. C., ... & Fischer, F. (2025). Personalizing simulation-based learning in higher education. *Learning and Individual Differences*, *122*, 102746.

* Giorgashvili, T., Jivet, I., Artelt, C., Biedermann, D., Bengs, D., Goldhammer, F., Hahnel, C., Mendzheritskaya, J., Mordel, J., Onofrei, M., Winter, M., Wolter, I., Horz, H., & Drachsler, H. (2025). From reflection to action: A controlled field study on how learners interpret and respond to feedback in learning analytics dashboards. *Journal of Computer Assisted Learning*, *41*(4).

*Jung, Y., & Wise, A. F. (2025). How students engage with learning analytics: Access, action-taking, and learning routines with message-based information to support collaborative annotation. *Computers & Education*, *232*, 105280.

*Küchemann, S., Avila, K. E., Dinc, Y., Hortmann, C., Revenga, N., Ruf, V., Stausberg, N., Steinert, S., Fischer, F., Fischer, M., Kasneci, E., Kasneci, G., Kuhr, T., Kutyniok, G., Malone, S., Sailer, M., Schmidt, A., Stadler, M., Weller, J., & Kuhn, J. (2025). On opportunities and challenges of large multimodal foundation models in education. NPJ Science of Learning.

Lave, J., & Wenger, E. (1991). Situated learning: legitimate peripheral participation. Cambridge University Press.

Lee, H. (2024). *Examining the effectiveness of personalized learning through artificial intelligence*. Indiana University.

Lee, H., Kim, H., & Yan, W. (2025). Promoting inclusive AI and technology in K-12: A review of context, instructional strategies, and learning outcomes. *Computers and Education: Artificial Intelligence*, 100478.

*Maier, U., & Klotz, C. (2025). Students ignore their mistakes: Elaborated error feedback processing in a digital learning system. Contemporary Educational Psychology, 82, Article 102395.

*Ning, J., Ma, Z., Yao, J., Wang, Q., & Zhang, B. (2025). Personalized learning supported by learning analytics: a systematic review of functions, pathways, and educational outcomes. Interactive Learning Environments, 1–23.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, Article n71.

*Tretow-Fish, T. A. B., & Khalid, Md. S. (2023). Methods for evaluating learning analytics and learning analytics dashboards in adaptive learning platforms: A systematic review. *Electronic Journal of E-Learning*, *21*(5), 430–449.

Tricco, A. C., et al. (2015). Scoping reviews. *Annals of Internal Medicine*.

Verbert, K., Ochoa, X., De Croon, R., Dourado, R. A., & De Laet, T. (2020). Learning analytics dashboards: the past, the present and the future. *Proceedings of the 10th International Conference on Learning Analytics & Knowledge*, 35–40.

*Weidlich, J., Fink, A., Frey, A., Jivet, I., Gombert, S., Menzel, L., Giorgashvili, T., Yau, J., & Drachsler, H. (2025). Highly informative feedback using learning analytics: how feedback literacy moderates student perceptions of feedback. *International Journal of Educational Technology in Higher Education*, *22*(1),

Wise, A. F., & Shaffer, D. W. (2023). Inclusive data governance. *Journal of Learning Analytics*.

*Wu, X.-Y., & Chiu, T. K. F. (2025). Integrating learner characteristics and generative AI affordances to enhance self-regulated learning: A configurational analysis. *Journal of New Approaches in Educational Research*, *14*(1), Article 10.

Yan, W., Lee, H., & Dancer, D. (2025, March). Assessing AI-Powered tools to teach data literacy in K-12. In *Society for Information Technology & Teacher Education International Conference* (pp. 3074-3080). Association for the Advancement of Computing in Education (AACE).

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory into Practice*, *41*(2), 64–70.

International Society of the Learning Sciences

Adaptive Support for Collaborative Learning: Effects of Design Features Across Task Types

Lixiang Gao, Frank Fischer, Olga Chernikova
Lixiang.Gao@campus.lmu.de, frank.fischer@psy.lmu.de, o.chernikova@psy.lmu.de
Ludwig-Maximilian-University (LMU), Munich

Abstract: This meta-analysis examines how adaptive instructional support facilitates collaborative learning across task types and design features, based on 80 studies. Cognitive-model approaches were most effective for factual knowledge learning, whereas data science approaches showed the strongest effects in problem-solving. Human-involved adaptation yielded smaller effects. These findings suggest adaptive instructional support is most effective when its design features align with task demands, with data science approaches and system control being particularly effective for problem-solving tasks.

Introduction and theoretical background

Collaborative learning (CL) supports diverse activities ranging from acquiring factual knowledge to jointly solving complex problems (Baker & Reimann, 2024). Adaptive instructional support aims to tailor learning environments to learners' needs but may not work equally across different task types. Prior research has often focused on a single task type or type of instructional support, offering limited insight into which types of instructional support are effective for which tasks in CL. This meta-analysis therefore aims to investigate how key design features address specific demands of different task types in CL.

Task types in CL can be distinguished into factual knowledge learning, focused on understanding and organizing domain knowledge, and problem-solving tasks involving the application of knowledge to address concrete problems (Lou et al., 2001; Pfister & Oehl, 2009). Because these tasks differ in complexity, regulation demands, and cognitive load, they might require different approaches to support.

One key design feature is the adaptative approach that indicates how support is adjusted to learners' needs (Plass & Pawar, 2020). Cognitive model approaches compare learner and expert models to identify knowledge gaps and improvement paths, making them effective for knowledge co-construction (Haq et al., 2021). Data science approaches such as learning analytics detect learners' needs from various learning data and adapt support (Plass & Pawar, 2020). Human-led adaptation provided by teachers or tutors benefits from their experience but may become demanding when multiple groups require timely support, especially in complex problem-solving tasks (Lund, 2004). A second design feature is decision authority, referring to whether learners, instructors, or systems regulate the amount, type, or timing of support (Kollar et al., 2018; Plass & Pawar, 2020). Learner control may enhance motivation by providing autonomy (e.g., Radkowitsch et al., 2021), but it requires learners to strategically apply meta-cognitive skills (Vandewaetere et al., 2011). Recent evidence suggests that system-controlled adaptation often yields stronger effects (Chernikova et al., 2025), especially in problem-solving tasks, where cognitive demands may limit learners' capacity to request support appropriately (Karich et al., 2014). To examine how task type, adaptive approach, and decision authority interact in shaping effectiveness, the following meta-analysis was conducted.

Research questions

This study aims to address the following research questions: (1) To what extent do task types moderate the effectiveness of adaptive instructional support in CL? (2) To what extent do approaches to adaptation moderate the effectiveness of instructional support across varied task types in CL? (3) To what extent does the decision authority in adaptation influence the effectiveness of instructional support across varied task types in CL?

Methods

Primary studies were retrieved from EBSCO, Web of Science, and Scopus using the search terms (adapt* OR computer* OR cognitive tutor OR intelligent tutoring system OR personalized) AND (scaffold* OR instruct*) AND (skill OR competenc* OR performance OR knowledge OR group learning) AND (CSCL OR computer supported collaborative learning OR collaborative learning OR cooperative learning) across all search fields. All records indexed before June 8, 2023 were considered, yielding 7,907 articles after duplicate removal. Following abstract and full-text screening, 80 studies met the inclusion criteria (adaptive intervention, CL context, and sufficient statistical data). Two coders independently coded all studies and resolved disagreements through

 © ISLS

discussion. Effects were estimated using Hedges' *g* within random-effects models (Borenstein et al., 2009). The analysis examined interactions between task types and design features of adaptive instructional support. Task type distinguished factual knowledge learning (e.g., concepts) from problem-solving. Adaptive approaches were coded as cognitive model (learner–expert model comparison for identifying knowledge gaps), data science (data-driven methods such as network analysis), or human support (teacher- or facilitator-provided support). Decision authority indicated who determined timing, amount, or type of adaptive support (learner, system, or teacher).

Results

Adaptive support appears slightly more effective for problem-solving than for factual knowledge learning (Table 1). For RQ2, the cognitive model approach yielded a large effect size (g = 1.01, SE = 0.31) for factual knowledge learning and a medium effect (g = 0.75, SE = 0.13) for problem-solving. Data science methods demonstrated a medium effect for factual knowledge learning (g = 0.69, SE = 0.13) and a large effect for problem-solving (g = 0.86, SE = 0.23). Human-led support showed small and inconsistent effects across task types (Table 1). For RQ3, system-controlled or teacher-controlled adaptive support yielded medium to high effects across tasks (g = 0.56 — 1.16, Table 2), while learner-controlled support showed smaller effects, with small to medium effects for problem-solving tasks (g = 0.33, SE = 0.15) and factual knowledge learning (g = 0.53, SE = 0.12).

Table 1

Interaction Effects of the Approach to Adaptation on CL with Different Task Types

	Overall	Cognitive Model	Data science	Human support
Factual knowledge learning	0.68^{**} (0.13) CI [0.41, 0.94] N = 32	1.01^{**} (0.31) CI [0.21, 1.8] N = 6	0.69^{**} (0.13) CI [0.42, 0.96] N = 14	0.76^{ns} (0.30) CI [-0.08, 1.61] N = 5
Problem-solving	0.76^{**} (0.11) CI [0.54, 0.99] N = 48	0.75^{**} (0.13) CI [0.42, 1.07] N = 7	0.86^{**} (0.23) CI [0.38, 1.35] N = 17	0.46^{**} (0.11) CI [0.21, 0.71] N = 11

Notes: Hedges' g (SE) reported; CI = confidence interval; ** $p < .05$; ^{ns}, not significant.

Table 2

Interaction Effects of the Decision Authority in Adaptation on CL with Different Task Types

	Learner	System	Teacher
Factual knowledge learning	0.53^{ns} (0.12) CI [-0.96, 2.03] N = 2	0.63^{**} (0.15) CI [0.32, 0.95] N = 19	1.16^{ns} (0.31) CI [-0.19, 2.51] N = 3
Problem-solving	0.33^{ns} (0.15) CI [-0.07, 0.73] N = 6	0.86^{**} (0.14) CI [0.57, 1.15] N = 27	0.56^{**} (0.19) CI [0.10, 0.12] N = 7

Notes: Hedges' g (SE) reported; CI = confidence interval; ** $p < .05$; ^{ns}, not significant.

Discussion

This meta-analysis confirms that adaptive instructional support facilitates CL in both knowledge co-construction tasks and problem-solving tasks, while its effectiveness varies depending on the interaction between task type and design features. Data science approaches showed particular potential for problem-solving tasks, consistent with previous findings (Tan et al., 2022). Cognitive model approaches were especially effective for factual knowledge tasks (Magnisalis et al., 2011). In contrast, human-led adaptation appeared less effective in problem-solving contexts, possibly because the complexity of group interaction places high demands on facilitators, making real-time diagnosis and regulation challenging (van Leeuwen & Janssen, 2019). Although human support may have strong potential to respond to interpersonal dynamics, its effectiveness appears limited in problem-solving tasks that require support for complex coordination and monitoring. Learner-controlled adaptation showed smaller effects, likely due to high cognitive load in problem-solving tasks, which leaves learners with fewer resources to adjust support (Vogel et al., 2022), despite its potential motivational benefits.

Overall, adaptive support appears most effective when its design features align with task demands. These findings highlight the importance of task-sensitive scaffolding. For example, data science approaches and system control may be particularly effective for problem-solving tasks. A limitation of this meta-analysis is that combinations of design features were not examined due to the limited number of studies. Future research should examine how specific configurations of adaptive approaches and decision authority differentially support factual knowledge and problem-solving tasks.

References

Baker, M. J., & Reimann, P. (2024). CSCL has come of age. *International Journal of Computer-Supported Collaborative Learning*, *19*(1), 1–10. https://doi.org/10.1007/s11412-024-09419-2

Borenstein, M., Hedges, L. V., Higgins, J., & Rothstein, H. (2009). *Introduction to meta-analysis*. John Wiley & Sons.

Chernikova, O., Sommerhoff, D., Stadler, M., Holzberger, D., Nickl, M., Seidel, T., Kasneci, E., Küchemann, S., Kuhn, J., Fischer, F., & Heitzmann, N. (2025). Personalization through adaptivity or adaptability? A meta-analysis on simulation-based learning in higher education. *1747-938X*, *46*, 100662. https://doi.org/10.1016/j.edurev.2024.100662

Haq, I. U., Anwar, A., Rehman, I. U., Asif, W., Sobnath, D., Sherazi, H. H. R., & Nasralla, M. M. (2021). Dynamic group formation with intelligent tutor collaborative learning: a novel approach for next generation collaboration. *IEEE ACCESS*, *9*, 143406–143422. https://doi.org/10.1109/ACCESS.2021.3120557

Karich, A. C., Burns, M. K., & Maki, K. E. (2014). Updated meta-analysis of learner control within educational technology. *Review of Educational Research*, *84*(3), 392–410. https://doi.org/10.3102/0034654314526064

Kollar, I., Wecker, C., & Fischer, F. (2018). Scaffolding and scripting (computer-supported) collaborative learning. In *International Handbook of the Learning Sciences* (pp. 340–350). Routledge. https://doi.org/10.4324/9781315617572-33

Lou, Y., Abrami, P. C., & d'Apollonia, S. (2001). Small group and individual learning with technology: A meta-analysis. *Review of Educational Research*, *71*(3), 449–521. https://doi.org/10.3102/00346543071003449

Lund, K. (2004). Human support in CSCL: what, for whom, and by whom? In *What we know about CSCL: And implementing it in higher education* (pp. 167–198). Springer.

Magnisalis, I., Demetriadis, S., & Karakostas, A. (2011). Adaptive and intelligent systems for collaborative learning support: A review of the field. *IEEE Transactions on Learning Technologies*, *4*(1), 5–20. https://doi.org/10.1109/tlt.2011.2

Pfister, H.-R., & Oehl, M. (2009). The impact of goal focus, task type and group size on synchronous net‐based collaborative learning discourses. *Journal of Computer Assisted Learning*, *25*(2), 161–176. https://doi.org/10.1111/j.1365-2729.2008.00287.x

Plass, J. L., & Pawar, S. (2020). Toward a taxonomy of adaptivity for learning. *Journal of Research on Technology in Education*, *52*(3), 275–300. https://doi.org/10.1080/15391523.2020.1719943

Radkowitsch, A., Sailer, M., Schmidmaier, R., Fischer, M. R., & Fischer, F. (2021). Learning to diagnose collaboratively – Effects of adaptive collaboration scripts in agent-based medical simulations. *Learning and Instruction*, *75.* https://doi.org/10.1016/j.learninstruc.2021.101487

Tan, S. C., Lee, A. V. Y., & Lee, M. (2022). A systematic review of artificial intelligence techniques for collaborative learning over the past two decades. *2666-920X*, *3*, 100097. https://doi.org/10.1016/j.caeai.2022.100097

van Leeuwen, A., & Janssen, J. (2019). A systematic review of teacher guidance during collaborative learning in primary and secondary education. *1747-938X*, *27*, 71–89. https://doi.org/10.1016/j.edurev.2019.02.001

Vandewaetere, M., Desmet, P., & Clarebout, G. (2011). The contribution of learner characteristics in the development of computer-based adaptive learning environments. *Computers in Human Behavior*, *27*(1), 118–130. https://doi.org/10.1016/j.chb.2010.07.038

Vogel, F., Kollar, I., Fischer, F., Reiss, K., & Ufer, S. (2022). Adaptable scaffolding of mathematical argumentation skills: The role of self-regulation when scaffolded with CSCL scripts and heuristic worked examples. *International Journal of Computer-Supported Collaborative Learning*, *17*(1), 39–64. https://doi.org/10.1007/s11412-022-09363-z

Acknowledgments

The research was performed with financial support by Deutsche Forschungsgemeinschaft, for TRR 419 SHARP and China Scholarship Council.

International Society of the Learning Sciences

Agentic Collaboration with GenAI: A Cautionary but Optimistic Tale

Iris Tabak, Ben-Gurion University of the Negev, itabak@bgu.ac.il
Gur Arie Livni Alcasid, Ben-Gurion University of the Negev, livnialc@post.bgu.ac.il
Ilana Dubovi, Tel-Aviv University, ilanadubovi@tauex.tau.ac.il

Abstract: Generative artificial intelligence (GenAI) extends human performance but may diminish users' sense of ownership. Drawing on research in CSCL and sociocultural theories of agency, we analyzed 559 naturalistic conversations from the LMSYS-Chat1M dataset, and coded for collaboration (e.g., Externalization, Co-construction) and agency (e.g., Critical Vigilance). Deep collaboration and high agency were infrequent but related. This suggests that CSCL research can inform the design of agency-supportive human-GenAI interactions.

Introduction

Generative artificial intelligence (GenAI) can extend people's capabilities to accomplish previously unattainable tasks, such as producing professional-quality videos or enabling non-programmers to write scripts. Yet, there is growing concern over diminished human agency (Favero, Pérez-Ortiz, Käser, & Oliver, 2025; Mei et al., 2025; Yepiskoposyan, 2025). Agency is the capacity to act to achieve one's goals through volition, control, ownership, and engagement. But if activity is increasingly offloaded to GenAI, then people may lose this sense of agency over the processes and products of user-GenAI interactions. We share this concern and aim to understand how we can maintain human agency. To this end, we analyzed conversations with GenAI chatbots. We draw on sociocultural approaches (Ahearn, 1999; Bakhtin, 1981; Wertsch, Tulviste, & Hagstrom, 1996) that consider agency as context-dependent and accomplished through mediational means. This implies that agency does not reside in the AI or the individual, but emerges from interactions among the setting, task, available resources, user, and GenAI. Our grounding assumption is that if interactions between humans and AI reflect deep collaboration akin to that experienced among humans (Roschelle & Teasley, 1995; Vogel, Weinberger, & Fischer, 2021), then such interactions may support agency. We asked: (1) Do conversations between users and GenAI chatbots reflect collaborative moves as previously identified in human-human collaboration? (2) Are there markers of user agency in user-GenAI conversations? (3) Is there a relationship between deep collaboration and high user agency? (4) Can we identify features of conversation that advance or impede user agency?

Methods

The data were drawn from an openly available corpus of naturalistic conversations with GenAI chatbots ("*assistant*") collected from April to August 2023 [https://huggingface.co/datasets/lmsys/lmsys-chat-1m] (Zheng et al., 2023). Anyone ("*user*") could access these chatbots freely and anonymously after consenting to release conversation data. To maximize the likelihood of observing collaboration, we included only problem-solving conversations (559 total). Table 1 shows a breakdown of turns and words. We adapted a collaboration coding scheme from CSCL research (Buseyne et al., 2024; Rosé et al., 2008; Weinberger & Fischer, 2006), (Table 2), and developed an agency scheme (Table 3) based on theoretical conceptualizations (Ahearn, 2001; Bakhtin, 1981; Wertsch et al., 1996). We expect that interactions that go beyond transactional to engage with or take up the other's contributions (Roschelle & Teasley, 1995) are more likely to be associated with agency. Therefore, we derived two new composite variables: *deep collaboration* and *high agency*. We used ChatGPT (OpenAI, 2025) to classify and filter the dataset and to code the conversation turns.

Table 1
Descriptive Statistics for User-Assistant Conversations

	Turns - minimum	Turns - maximum	Turns - median	Words - minimum	Words - maximum	Words - median
Assistant	2	11	2	2	2129	379
User	2	11	2	4	1197	30

Table 2
Collaboration Moves Coding Scheme Applied to User and Assistant

Code	Definition
A1 Externalization	Partner contributes information without reference to prior discussion.
A2 Elicitation	One partner explicitly seeks input, ideas, or elaboration from the other.

 © ISLS

A3 Quick Consensus	A partner agrees without substantive engagement.
A4 Integration	A partner incorporates or modifies their own idea using the other's contribution.
A5 Conflict Co-construction	A partner challenges or contrasts ideas to refine shared understanding.

Table 3

Agency Markers Coding Scheme Applied to User and Assistant

Code	Definition
Self-initiated goals	Propose goals, reframe tasks.
Critical vigilance	Monitor content (appraisal, seek evidence, compare to other sources)
Constructive trust	Trusting other with stipulations, integrating other's perspective with one's own.
Blind acceptance	[Negative marker] Take up other's perspective without evaluation.
Resistance	Challenge, correct, modify other's perspective.
Voice	Taking on an authorial voice and positioning oneself as equal and partner to the other.

Findings

We observed a range of collaborative and agency moves (Tables 4 and 5). There were no clear trends, and collaboration was mostly transactional. Although deep collaboration, high agency, and their intersection (21 instances) were infrequent, there was a significant relationship between deep collaboration and high agency $\chi^2(1, N = 559) = 7.23$, $p = .007$. We were unable to find significant predictors of high agency.

Table 4

Frequency of Collaboration Moves by User and Assistant

	A1 Externalization	**A2** Elicitation	**A3** Consensus	**A4** Integration	**A5** Conflict	
Assistant	708	45	80	124	541	1498
User	1076	309	20	15	85	1505
	1784	354	100	139	626	

Table 5

Frequency of Agentic Markers by User and Assistant

	Self-initiated goals	Critical vigilance	Constructive trust	Blind acceptance	Resistance	Voice
Assistant	161	158	29	102	52	**120**
User	517	512	98	395	335	**236**

Discussion

As the use of GenAI grows, so does the concern that it will erode human agency. If, as we posit, agency is a function of available resources, settings, and social positioning, then the qualities of the interaction could play a role in sustaining or dampening agency. Therefore, we examined whether human-GenAI conversations reflect agentive and collaborative moves. Deep collaboration and high agency were rare but correlated, motivating further research. We need to uncover the prevalence, strength, and direction of this relationship. We also need to identify whether individual practices and dispositions, such as intellectual perseverance, can counter the tendency to defer intellectual engagement to machines (Turner, 2022). Consistent with our findings, recent work calls for the design of shared human-machine agency (Mathewson et al., 2023). Prior research in CSCL and the Learning Sciences, such as research on how to design for a balance between authoritative and internally persuasive voice (Tabak, 2020; Tabak & Baumgartner, 2004), and on how to guide and structure deep collaboration (Ludvigsen, Cress, Law, Rosé, & Stahl, 2016; Rummel, Spada, & Hauser, 2009; Vogel, Weinberger, & Fischer, 2021), can inform these efforts.

Conclusions and significance

One can hardly go through a day without encountering some mention of GenAI. Much of the discourse revolves around risks, such as the displacement of learning and the dissemination of inaccurate information. This study, alongside others (Favero et al., 2025; Mei et al., 2025; Yepiskoposyan, 2025), extends the conversation to issues of diminishing agency. Our findings tell a cautionary but optimistic tale. Sustaining human agency is paramount because even if technological advances eliminate GenAI errors, the need for agency remains. If people lose the essence of what it is to be human, these advancements will have been for naught.

References

Ahearn, L. M. (1999). Agency. *Journal of Linguistic Anthropology, 9*(1/2), 12-15.

Ahearn, L. M. (2001). Language and agency. *Annual Review of Anthropology, 30*(1), 109-137.

Bakhtin, M. M. (1981). *The dialogic imagination: Four Essays by M. M. Bakhtin* (C. Emerson & M. Holquist, Trans.). Austin, TX: University of Texas Press.

Buseyne, S., Rajagopal, K., Danquigny, T., Depaepe, F., Heutte, J., & Raes, A. (2024). Assessing verbal interaction of adult learners in computer-supported collaborative problem solving. *British Journal of Educational Technology, 55*(4), 1465-1485.

Favero, L., Pérez-Ortiz, J.-A., Käser, T., & Oliver, N. (2025). Do AI tutors empower or enslave learners? Toward a critical use of AI in education. *arXiv preprint arXiv:2507.06878*.

Ludvigsen, S., Cress, U., Law, N., Rosé, C. P., & Stahl, G. (2016). Collaboration scripts and scaffolding. *International Journal of Computer-Supported Collaborative Learning, 11*(4), 381-385. doi:10.1007/s11412-016-9247-1.

Mathewson, K. W., Parker, A. S. R., Sherstan, C., Edwards, A. L., Sutton, R. S., & Pilarski, P. M. (2023). Communicative capital: A key resource for human–machine shared agency and collaborative capacity. *Neural Computing and Applications, 35*(23), 16805-16819. doi:10.1007/s00521-022-07948-1.

Mei, P., Brewis, D. N., Nwaiwu, F., Sumanathilaka, D., Alva-Manchego, F., & Demaree-Cotton, J. (2025). If ChatGPT can do it, where is my creativity? Generative AI boosts performance but diminishes experience in creative writing. *Computers in Human Behavior: Artificial Humans, 4*, 100140.

OpenAI. (2025). ChatGPT (GPT-5) [large language model]. Retrieved from https://chatgpt.com

Roschelle, J., & Teasley, S. (1995). The construction of shared knowledge in collaborative problem solving. In C. O'Malley (Ed.), *Computer Supported Collaborative Learning* (CSCL) (pp. 69–97). Berlin: Springer-Verlag.

Rosé, C., Wang, Y.-C., Cui, Y., Arguello, J., Stegmann, K., Weinberger, A., & Fischer, F. (2008). Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in computer-supported collaborative learning. *International Journal of Computer-Supported Collaborative Learning, 3*(3), 237-271. doi:10.1007/s11412-007-9034-0

Rummel, N., Spada, H., & Hauser, S. (2009). Learning to collaborate while being scripted or by observing a model. *International Journal of Computer-Supported Collaborative Learning, 4*(1), 69-92.

Tabak, I. (2020). Post-truth GPS: Detour at truth, take the long route to useful knowledge. *Educational Psychologist, 55*(3), 181-186. doi:10.1080/00461520.2020.1784734

Tabak, I., & Baumgartner, E. (2004). The teacher as partner: Exploring participant structures, symmetry, and identity work in scaffolding. *Cognition and Instruction, 22*(4), 393-429.

Turner, C. (2022). Neuromedia, cognitive offloading, and intellectual perseverance. *Synthese, 200*(2), 66.

Vogel, F., Weinberger, A., & Fischer, F. (2021). Collaboration scripts: Guiding, internalizing, and adapting. In *International handbook of computer-supported collaborative learning* (pp. 335-352): Springer.

Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & Education, 46*(1), 71-95.

Wertsch, J. V., Tulviste, P., & Hagstrom, F. (1996). A sociocultural approach to agency. In E. A. Forman, N. Minick, & C. A. Stone (Eds.), *Contexts for learning : Sociocultural dynamics in children's development*. Cary, USA: Oxford University Press, Incorporated.

Yepiskoposyan, R. (2025). How perceived agency in Generative AI undermines critical thought. *Available at SSRN 5603171*.

Zheng, L., Chiang, W.-L., Sheng, Y., Li, T., Zhuang, S., Wu, Z., . . . Xing, E. P. (2023). Lmsys-chat-1m: A large-scale real-world LLM conversation dataset. *arXiv preprint arXiv:2309.11998*.

Acknowledgments

Funded in part by the Ministry of Science and Culture of Lower Saxony and the Volkswagen Foundation as part of the Israel-Lower Saxony cooperation grant to Hendriks, Thies, Dubovi & Tabak for the EngageAI project.

International Society of the Learning Sciences

Sociotechnical Efficacy in Modeling 3D Interactive Environments

Akin Adediji, Jinhee Lee
akin.adediji@pepperdine.edu, jinhee.lee2@pepperdine.edu
Pepperdine University

Abstract: This study examines sociotechnical efficacy in 3D environment modeling through the Roblox Developers Forum. Ito & Reich's (2017) equity-centered principles explore how youth learning participation is fostered in creative spaces. Available curriculum, plugins, and developer feedback built around "hanging out", "messing around", "geeking out" (HOMAGO) emphasize the "hanging out" dimension (Ito, 2009). Platforms enhanced by modular, adaptive assets foster equitable access and social connectedness, shifting learning between participatory HOMAGO to supporting agency and connectivist models.

Introduction

This study examines how developer forum affordance's structure participation and learning within environments focused on youth-based sociotechnical efficacy. We analyze contemporary sociotechnical patterns in forum commentary to understand user satisfaction with participation. The discussion associate's constructionist technological design with connectivist learning theory (Harasim, 2017). It frames Roblox as a desirable platform that serves as both a creative metaverse and a social medium.

As Roblox Studio incorporates generative AI technologies for mainstream 3D development, youth engagement in computer-mediated environments increasingly blends creation, collaboration, and play. The HOMAGO framework situates this engagement spectrum within distinct learning dispositions (Ito et al., 2019). Rather than isolated innovation, youth creators engage in iterative, modular technical feedback loops with peers, enhancing agency (constructionist making) and connection (connectivist learning) (Peppler & Thompson, 2024). Generative AI assets amplify this dynamic by simplifying creation, freeing cognitive space for relational and aesthetic collaboration. This grounded study proposes the research question: in what ways do generative AI tools and modular 3D environment design features in Roblox Studio influence literacy development in youth-centered digital communities?

Theoretical framework

Ito et al.'s (2019) HOMAGO study remains foundational for understanding and documenting how youth learn through interest-driven participation in digital media. Insight into "Equity by Design" calls for learning technologies that align home, school, and community contexts to bridge socio-technical divides (Ito & Reich, 2017). Together, these frameworks support a sociotechnical lens that values open participation and access. Roblox's plugin-based systems illustrate Ito and Reich's (2017) principle that 'open ≠ equitable.' The platforms community-driven structure enables learners to co-develop technical literacies; however, it implies that it will increase accessibility for all over time. While the tools are technically available for free, equitable learning arises only when their use connects to youth interests, identities, and peer mentorship.

Case context

This poster presents findings on Roblox's Updates & Community Events DevForum post "2025 Year in Review!". The post highlighted new tools introduced to Roblox over the last year including: custom emotes, server authority/ban API, events detail page, emissive maps, speech-to-text API, text-to-speech API, and physically-based rendering for accessories. These tools lower barriers to participation and create conditions for peer-to-peer engagement, described as "geeking out" mode, where learning is distributed across design communities and individual experimentation is reinforced. Findings suggest that modular embedded design tools, like plugins, transform 'messing around' behaviors into collective "geeking out" practices.

Methods

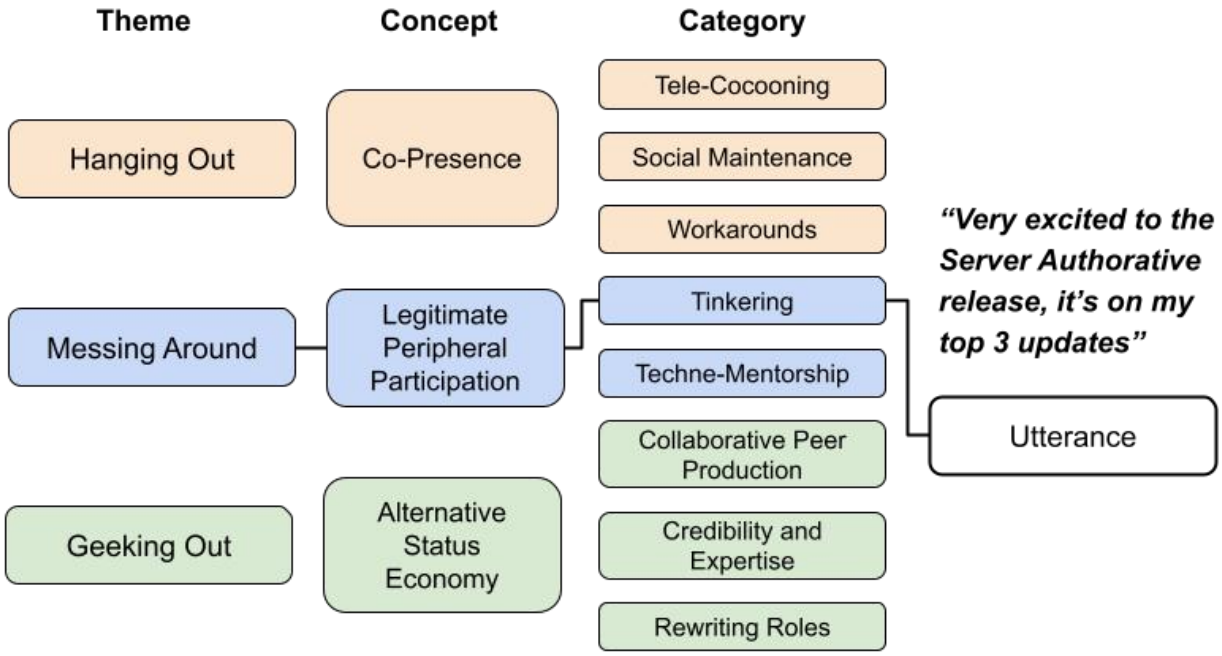
This study will employ thematic analysis of developer comments on the Roblox DevForum. Using a deductive approach, *a priori* themes aligned with HOMAGO categories guide analysis by identifying patterns of sociotechnical affordances in 3D development workflows. Each utterance is coded, then placed in category that aligns with an overarching HOMAGO theme. Of 123 comments, 113 utterances from the post were analyzed using thematic analysis to examine how the seven tools transform technical literacy practices into socially meaningful interactions. The dataset excluded direct replies to comments and any images included in posts.

 © ISLS

Figure 1
Experience Details Page Video and Events Analytics



Figure 2
Example of Coded Utterance



Findings

Preliminary findings suggest that rather than isolated innovation, youth creators engage in iterative, modular technical feedback loops with peers, enhancing agency (constructionist making) and connection (connectivist learning) (Peppler & Thompson, 2024). The analysis of 113 developer utterances provides empirical support for this dynamic, as the high frequency of Social Maintenance *(n* = 25) and Tele-Cocooning (*n* = 20) demonstrates that "hanging out" serves as a vital precursor to technical collaboration. Generative AI assets amplify this dynamic by simplifying creation, freeing cognitive space for relational and aesthetic collaboration. This is further evidenced by the prevalence of Tinkering *(n* = 16) and Techne-Mentorship (*n = 12*), in which participants use modular tools to engage in higher-level "geeking out" behaviors.

Implications

The social interaction sample highlights the need to assess equity in technological cultures rather than assuming open-source ensures inclusion. Emerging generative AI systems within creative 3D platforms like Roblox risk reinforcing hierarchical participation, where technical proficiency can diminish collaborative engagement, leading to reduced motivation and negative environmental affordances for users, and a larger technology gap. Practices like co-creation and other accessible forms of 'hanging out' should be leveraged as inclusive practices that extend beyond isolated production toward shared experience. Sociotechnical efficacy depends not only on individual mastery but collective equity in connection and creation.

 © ISLS

References

Harasim, L. M. (2017). *Learning theory and online technologies* (2nd ed.). Routledge, Taylor & Francis Group.

Ito, M., Baumer, S., Bittanti, M., Boyd, D., Cody, R., Herr-Stephenson, B., et al. (2019). *Hanging out, messing around, and geeking out: Kids living and learning with new media* (10th Anniversary ed.). MIT Press.

Ito, M., & Reich, J. (2017). *From good intentions to real outcomes: Equity by design in learning technologies*. Digital Media + Learning Research Hub.

Ozamanthus. (2025, December 19). 2025 year in review! Roblox Developer Forum. https://devforum.roblox.com/t/2025-year-in-review/4164479

Peppler, K. A., Thompson, N., Danish, J., & Moczek, A. (2018). In the hive: Designing for emergence when teaching complex systems in early childhood. In J. Kay & R. Luckin (Eds.), *Rethinking learning in the digital age: Making the Learning Sciences count, 13th International Conference of the Learning Sciences (ICLS) 2018* (Vol. 1, pp. 512–519). International Society of the Learning Sciences.

Peppler, K., & Thompson, N. (2024). Tools and materials as non-neutral actors in STEAM education. *Journal of the Learning Sciences, 33*(4–5), 719–756. DOI: 10.1080/10508406.2024.2380694

 © ISLS

International Society of the Learning Sciences

Designing an AI-Augmented Knowledge Synthesis Tool for Collaborative Learning

Xinran Zhu, Bo Shui, Ishita Asnani, Liam Magee
xrzhu@illinois.edu, boshui2@illinois.edu, iasnani2@illinois.edu, lmagee@illinois.edu
University of Illinois Urbana-Champaign

Abstract: Knowledge synthesis involves the deliberate integration of diverse information to generate new insights and creative solutions. In CSCL, it is both a process—where students advance collaborative knowledge—and a product that facilitates ongoing learning. We introduce a novel technology that scaffolds students' knowledge synthesis in collaborative learning environments augmented by generative AI. The poster presents its design rationale, technical details, and future directions for iterative development and classroom implementation to support knowledge synthesis in CSCL.

Introduction

Knowledge synthesis refers to the deliberate and strategic integration of diverse strands of information to foster conceptual innovation and design creative solutions (Deschryver, 2014; Morabito & Chan, 2021; Qian et al., 2020). It is a higher-order cognitive process that extends beyond remembering and understanding (Bloom et al., 1956), playing a central role in human learning and knowledge production. Broadly, knowledge synthesis drives conceptual change in science learning (Chi et al., 1994), fuels innovative scientific discoveries (Morabito & Chan, 2021), and sustains knowledge building processes that enable learners to rise above peers' ideas to advance shared understanding and foster innovation within communities (Scardamalia & Bereiter, 2014).

Knowledge synthesis plays a vital role in collaborative learning. Consider an undergraduate class where students engage in online discussions about their readings before coming to class. By synthesizing these discussions, students analyze and build on their collaborative knowledge, and the resulting synthesis artifacts help guide in-class dialogue to further deepen their thinking. In this sense, knowledge synthesis serves both as a *product* that reflects what students have learned and as a *mediator* that supports ongoing learning (Zhu et al., 2026). The quality (e.g., clarity, depth) and format (e.g., visual representation of ideas) of these synthesis artifacts are associated with how effectively they support ongoing learning. Effective synthesis artifacts not only capture key ideas but also make them easier to revisit, build upon, act on, and share in future collaborative tasks (Morabito & Chan, 2021; Zhu et al., 2026). However, creating effective syntheses is both rare and challenging due to the cognitive demands involved as well as inadequate support from information systems (Morabito & Chan, 2021; Qian et al., 2020; Zhu et al., 2026). In our previous study, the synthesis artifacts created by students tended to lack cognitive complexity and contextual information, limiting their potential for reuse and further evolution in computer-supported collaborative learning (CSCL) (Zhu et al., 2026).

To address these challenges, we present *Synthesis AI Lab* (SAIL), an AI-driven learning platform informed by CSCL and HCI research (e.g., Morabito & Chan, 2021; Qian et al., 2020; Scardamalia & Bereiter, 2014; Tricco et al., 2011). SAIL is designed to scaffold the complex cognitive work involved in synthesis. It helps students connect ideas, include sufficient contextual information, and produce synthesis artifacts that are actionable and generative for future learning. In our context, SAIL is designed for a CSCL environment that begins with social annotation. Social annotation refers to the practice of reading shared texts on digital platforms such as Perusall, where students highlight passages, write annotations, and respond to one another. This process makes students' emerging ideas visible and creates a foundation for collaborative learning. SAIL then supports students in synthesizing these ideas into more integrated forms that can inform later in-class discussions, group work, or writing tasks.

The innovation: Synthesis AI Lab (SAIL)

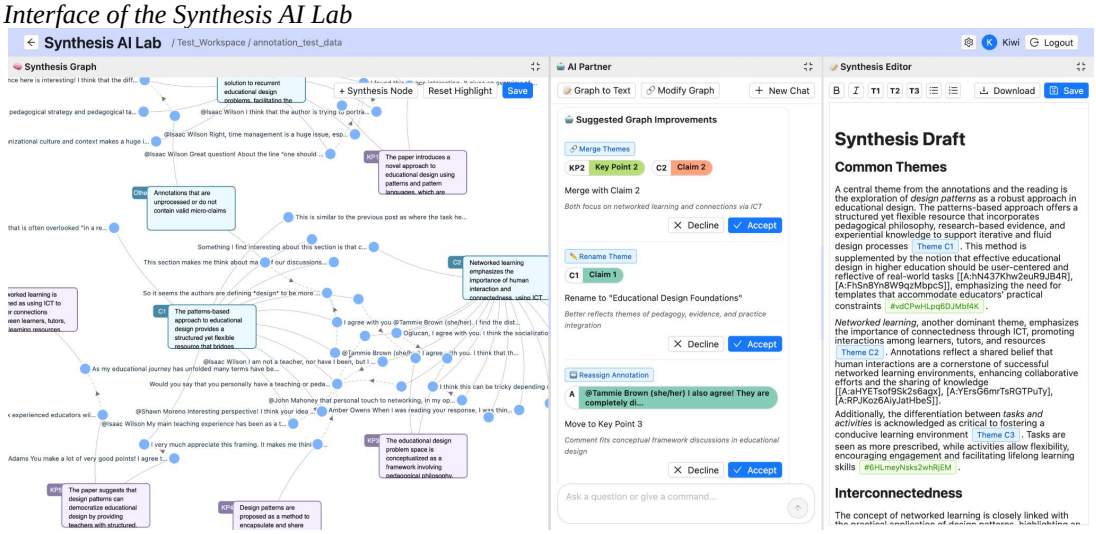
SAIL is an extension to the Synthesis Lab, which is a web-based application developed by our team to support students in synthesizing their contributions from social annotation activities (Zhu et al., 2023). It breaks down the complex process of synthesizing knowledge into smaller, manageable steps, such as distilling ideas from annotations and categorizing peers' ideas into conceptual building blocks. SAIL extends the Synthesis Lab in three ways, as shown in Figure 1. First, it introduces a *Knowledge Synthesis Graph* (Shui & Zhu, 2026) that visualizes the current state of student knowledge reflected in their annotations as a graph (left panel). Second, it introduces an *AI partner* that scaffolds the knowledge synthesis process by suggesting thematic links, surfacing

 © ISLS

conceptual gaps, and prompting reflection to engage in productive knowledge building activities (middle panel). Third, a *Synthesis Editor* is integrated with the graph and AI partner, enabling the automatic generation of text-based syntheses from the graph, which students can further edit (right panel).

As illustrated in Figure 1, the graph is structured around *individual annotations* (nodes), *synthesis nodes* (rectangles) representing higher-level ideas synthesized from related annotations, *epistemic relations* linking these elements, and metadata such as reply threads and author information. The graph is generated using large language models (LLMs). In the current prototype, GPT-4o (Hurst et al., 2024) was prompted at a low temperature (0.2).

Figure 1
Interface of the Synthesis AI Lab



SAIL's AI system generates an initial synthesis graph by processing data extracted from the social annotation platform Perusall. This graph then serves as a knowledge artifact for student interaction and refinement. Students can further analyze annotations by reassigning them to different synthesis nodes, create new synthesis nodes, revise existing ones, and refine the epistemic relations. They can also document their reasoning as contextual information for later use. Throughout this process, the AI Partner can support students in advancing the graph. It is powered by multiple AI agents that interact dynamically with the synthesis graph and annotation data via structured JSON files. These files encode the key elements of the graph, providing context for the LLMs. Working in coordination with the graph and the synthesis editor, the agents support the synthesis process as a form of human-AI collaboration. All the agents and workflows are incorporated with the Chain-of-Thought (Wei et al., 2022) technique to enhance LLM in reasoning capability and context information retrieval from the graph.

Three core features of the AI partner include: (1) In-Sync Interaction: As users engage with the graph, such as by clicking on a node, the system automatically highlights related elements (e.g., annotations, synthesis nodes, and linked replies). The AI partner provides summaries and follow-up prompts in the chat interface, encouraging focused and deeper thinking on the highlighted threads. (2) Reflection Promotion: Beyond manual graph interaction, the Modify Graph agent suggests actionable improvements. Users can accept or reject these suggestions or continue the dialogue to reflect and extend their thinking during the knowledge building process. (3) Connected Synthesis: The Graph-to-Text agent generates written syntheses from the synthesis graph and AI interactions. The synthesis text is linked to both the graph and the AI partner through automatic references to source annotations, making idea development traceable and preserving important contextual information. The resulting synthesis—a written document complementing the graph view—serves as an alternative representation of collaborative knowledge. Students iteratively refine this document to articulate connections, tensions, and emerging ideas from the annotations, as well as actionable next steps for their ongoing collaborative work.

Next steps

Future work will iteratively evaluate and refine the SAIL platform through design-based research. One example use case of SAIL is to help students synthesize annotations before class and use these syntheses to support in-class discussion. Classroom implementation studies will examine how the system supports human-AI collaboration in CSCL, offering empirical insights into AI-augmented learning and collaboration.

 © ISLS

References

Bloom, B. S. (1956). T*axonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain*. David McKay.

Brown, A. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *The Journal of Learning Sciences, 2*(2), 141–178. https://doi.org/10.1207/s15327809jls0202_2

Chi, M. T. H., Slotta, J. D., & De Leeuw, N. (1994). From things to processes: A theory of conceptual change for learning science concepts. *Learning and Instruction, 4*(1), 27–43. https://doi.org/10.1016/0959-4752(94)90017-5

Deschryver, M. (2014). Higher Order Thinking in an Online World: Toward a Theory of Web-Mediated Knowledge Synthesis. *Teachers College Record, 116*(12), 1–44. https://doi.org/10.1177/016146811411601202

Fischer, F., Bruhn, J., Gräsel, C., & Mandl, H. (2002). Fostering collaborative knowledge construction with visualization tools. *Learning and Instruction, 12*(2), 213–232. https://doi.org/10.1016/S0959-4752(01)00005-6

Hatano, G., & Inagaki, K. (1991). Sharing cognition through collective comprehension activity. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 331–348). American Psychological Association. https://doi.org/10.1037/10096-014

Hawkins, J., & Pea, R. D. (1987). Tools for bridging the cultures of everyday and scientific thinking. *Journal for Research in Science Teaching, 24*, 291–307. https://doi.org/10.1002/tea.3660240404

Hurst, A., Lerer, A., Goucher, A. P., Perelman, A., Ramesh, A., Clark, A., ... & Kivlichan, I. (2024). Gpt-4o system card. *arXiv preprint arXiv:2410.21276*.

Lave, J. (1987). *Cognition in practice*. Cambridge University Press.

Morabito, J. S., & Chan, J. (2021). Managing context during scholarly knowledge synthesis: Process patterns and system mechanics. *Creativity and Cognition*, 1–5. https://doi.org/10.1145/3450741.3465244

Qian, X., Fenlon, K., Lutters, W., & Chan, J. (2020). Opening up the black box of scholarly synthesis: Intermediate products, processes, and tools. *Proceedings of the Association for Information Science and Technology, 57*(1), e270. https://doi.org/10.1002/pra2.270

Scardamalia, M., & Bereiter, C. (2014). Knowledge Building and Knowledge Creation: Theory, Pedagogy, and Technology. In R. K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (2nd ed., pp. 397–417). Cambridge University Press. https://doi.org/10.1017/CBO9781139519526.025

Shui, B., & Zhu, X. (2026). Knowledge Synthesis Graph: An LLM-Based Approach for Modeling Student Collaborative Discourse. *arXiv*. https://doi.org/10.48550/arXiv.2602.06194

Tricco, A. C., Tetzlaff, J., & Moher, D. (2011). The art and science of knowledge synthesis. *Journal of Clinical Epidemiology, 64*(1), 11–20. https://doi.org/10.1016/j.jclinepi.2009.11.007

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., ... & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, *35*, 24824-24837.

Zhu, X., Chen, B., & DeLiema, D. (2026). Advancing collaborative discourse through knowledge synthesis. *International Journal of Computer-Supported Collaborative Learning*, 1-31. https://doi.org/10.1007/s11412-025-09463-6

Zhu, X., Shui, H., & Chen, B. (2023). The synthesis lab: Empowering collaborative learning in higher education through knowledge synthesis. In *Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing* (pp. 245-248). https://doi.org/10.1145/3584931.3606996

International Society of the Learning Sciences

Encouraging Deeper Learning with GenAI: A Review of Pedagogical Constraints on GenAI and Their Impact in Higher Education

Shuang Geng, Institute for Excellence in Teaching & Learning, Boston University, sgeng@bu.edu
Di Fan, University of Iowa, di-fan@uiowa.edu
Juhong Eom, Harvard Graduate School of Education, juhong_eom@gse.harvard.edu
Gahyun Callie Sung, University of Iowa, gahyun-sung@uiowa.edu

Abstract: This paper reviews how generative AI (GenAI) can be pedagogically constrained to promote deeper learning in higher education. Analyzing 15 intervention studies comparing constrained GenAI use with free or no GenAI use, we found three delivery methods (process, activity, and tool levels), operationalized through five mechanisms. Results show primarily positive impacts on performance, motivation, and higher order thinking, though nearly half rely on self-reports. The findings highlight the role of constraints in GenAI use, and invite further research explicitly linking constraints to outcomes.

Setting

There has been rapidly growing interest in finding ways to integrate generative AI (GenAI) into higher education classrooms. Meanwhile, there is little clarity on *how* GenAI use can be moderated to better facilitate learning. Existing reviews have either explored strategies for GenAI use or evaluated the impact of having GenAI in learning as a whole, but have yet to distinguish the specific moderating or constraint mechanisms and their impact on learning (Batista et al., 2024; Almasri, 2024). To bridge this gap, our review goes beyond a binary view (use versus no use) with the following guiding questions:

- What student-facing constraint strategies for GenAI have been tested in higher education?
- What impact has been observed so far, on what types of learning outcomes?

The current paper presents preliminary findings from an ongoing systematic review. We screened journal articles from 2022 (i.e., release of ChatGPT) to 2025, following the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) standard (Page et al., 2021). To be included, studies had to 1) investigate the impact of GenAI in higher education learning; 2) design and clearly describe a constrained or moderated GenAI use condition; 3) employ a (quasi-)experimental design, where academic or cognitive outcomes were compared between a treatment group of constrained use and a control group of free or no use of GenAI. Systematic search across Scopus, Web of Science, and ACM yielded 1,910 articles, among which 30 met the inclusion criteria. The present analysis reports findings based on half (15) of the final dataset, selected alphabetically by the first author's last name (see full list in References). All studies were collaboratively coded by the first author and at least one other author, with discrepancies resolved through discussion for this preliminary analysis.

Results

Context of Studies: Interventions showed a balanced representation between eight in STEM and seven in non-STEM disciplines. ChatGPT (or its API) was the primary tool in 14 of 15 cases.

Constraint Delivery: Our coding process revealed three main GenAI delivery methods. First, *process-level guidelines* (n=10) explicitized expectations for use, by delineating permissible forms of assistance (e.g., guiding usage with metacognitive support prompts; Xu et al., 2025). On the other hand, *activity-level scaffolds* (n=10) embedded constraints within the design of a learning activity, specifying the stages, timing, and conditions under which GenAI may be used (e.g., requiring independent initial attempt, followed by interaction and revision with GenAI; Jiang & Nakatani, 2025). Lastly, *tool-level constraints* (n=11) configured tools for specific interaction modes, such as limiting outputs to hints (Lee et al., 2024).

Constraint Mechanisms: Across interventions, five mechanisms emerged across the four stages of problem-solving: understanding, planning, execution, and revision (Polya, 1945). One mechanism constrained GenAI use to the early stage of *draft or brainstorming*, with the expectation that these ideas would be refined by the learner. The second and third mechanisms applied to GenAI use in the execution stage: *solution first* required students to create their own solution first, while *no answers* primed GenAI to not give out whole answers and instead provide Socratic guidance. Lastly, in the revision stage, mechanisms included *revision and verification*,

 © ISLS

as well as *attributing and reflecting*, where learners were either required to revise and verify AI output, or required to reflect on, and potentially report, how they used GenAI to perform the task.

A single intervention often had multiple constraint delivery methods and/or mechanisms, reflecting layered approaches rather than reliance on a single lever. Overall, the pattern indicates a shift from merely restricting access to deliberately structuring how GenAI participates in learning, with mechanisms intuitively chosen to externalize cognition, require evidence-seeking, and align tool behavior with instructional intent.

Impact on student learning

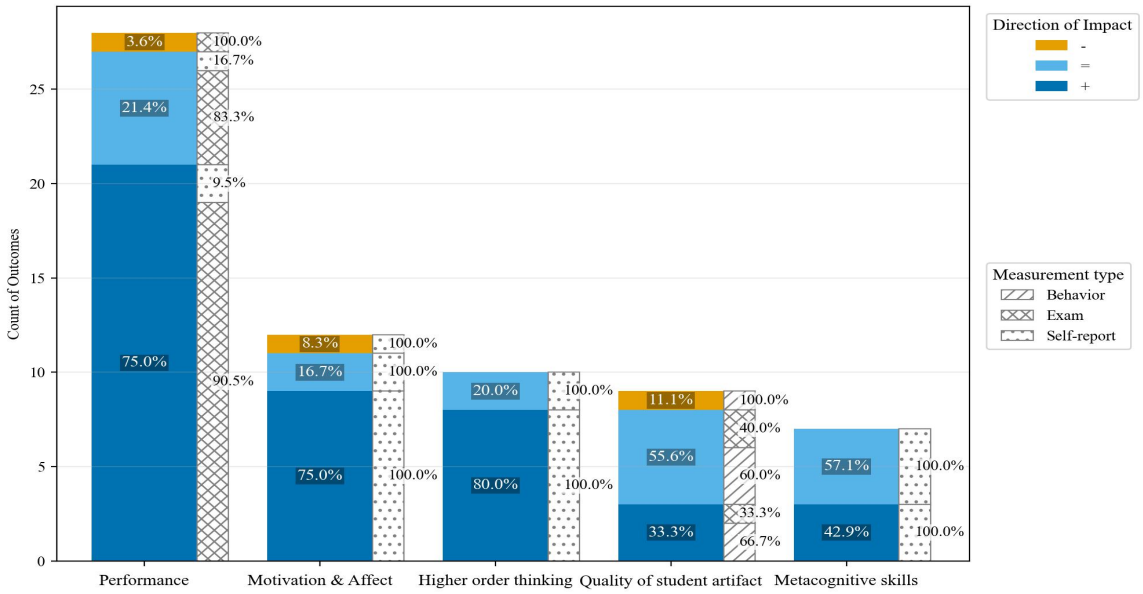
Learning Outcome: On average, 4.4 outcomes were measured per study. The most common outcome category was *Performance* (e.g., exam grade, skill evaluation score), totaling 28 outcome measures across 13 (86.7%) studies. Other outcome categories included Motivation and Affect (e.g., intrinsic motivation and engagement), Higher Order Thinking (e.g., critical thinking), Quality of Student Artifact, and Metacognitive Skills.

Direction of impact: Treatment groups showed largely positive impacts across student learning outcomes compared to control groups: 75% positive on Performance, 75% on Motivation & Affect, 80% on Higher Order Thinking, and the remaining two categories showed more mixed results. Figure 1 shows the breakdown of directions of impact for each outcome category, combined with a second-level breakdown of measurement type.

Measurement Type: Nearly half of the 66 outcomes (48.5%) relied on self-reports, which may have led to positive bias. Notably, outcomes related to *Higher Order Thinking* and *Motivation & Affect*, while showing predominantly positive results, relied exclusively on self-report (see Figure 1).

Figure 1

Outcome Categories by Direction of Impact and Measurement Type



Main bars: Count and percentage breakdown of direction of impact for each outcome category (X-axis). Side bars with hatching patterns: Percentage breakdown of measurement type for each main bar segment.

Discussion

Empirical findings point to a generally positive impact of constrained GenAI use on student performance, with more mixed results for metacognitive skills or quality of work. However, this direction may be influenced by several biases. In addition to common sources of bias such as the observer effect, limitations of self-reports, or the novelty effect, a unique issue was that in some cases, assessments were made on the GenAI-assisted task itself. Students may have shown higher performance because they had assistance, and not because they truly learned more. Another limitation was that most of these results were gained from comparing constrained GenAI use to no GenAI use. There was a surprising dearth of research (20%) that compared free versus constrained use of GenAI (c.f. comparing no and constrained use of GenAI), which could help provide stronger evidence of pedagogical value by isolating the effects of constraint design rather than GenAI presence alone. Our ongoing analysis of the full dataset aims to delineate such emerging design insights with greater granularity.

 © ISLS

References

Studies included in the review are marked with an asterisk (*).

Almasri, F. (2024). Exploring the impact of artificial intelligence in teaching and learning of science: A systematic review of empirical research. *Research in Science Education*, 54(5), 977-997.

*Atta, A., Esmat, M., Amasha, N., Elayat, E., & ElSaid, W. K. (2025). A smart ChatGPT mobile application for improving C# programming skills for students in educational institutions. *Qubahan Academic Journal*, *5*(2), 49–62.

Batista, J., Mesquita, A., & Carnaz, G. (2024). Generative AI and higher education: Trends, challenges, and future directions from a systematic literature review. *Information*, 15(11), 676.

*Chan, S. T. S., Lo, N. P. K., & Wong, A. M. H. (2024). Enhancing university level English proficiency with generative AI: Empirical insights into automated feedback and learning outcomes. *Contemporary Educational Technology, 16*(4), ep541.

*Fardian, D., Suryadi, D., Prabawanto, S., & Jupri, A. (2025). Integrating Chat-GPT in the classroom: A study on linear algebra learning in higher education. *International Journal of Information and Education Technology, 15*(4), 732-751.

*George, A., Storey, V. C., & Hong, S. (2025). Unraveling the impact of ChatGPT as a knowledge anchor in business education. *ACM Transactions on Management Information Systems, 16*(1), 1-30.

*Hu, Y., Chen, J., & Hwang, G. J. (2025). A ChatGPT-supported QIVE model to enhance college students' learning performance, problem-solving and self-efficacy in art appreciation. *Interactive Learning Environments*, 1-14.

*Huang, C., Zhong, Y., Li, Y., Wang, X., Han, Z., Zhang, D., & Liu, M. (2025). Enhancing student reading performance through a personalized two-tier problem-based learning approach with generative artificial intelligence. *Humanities and Social Sciences Communications*, *12*(1), 1-16.

*Huang, Y. M., Chen P. H., Lee, H. Y., Sandnes, F. E., & Wu, T. T. (2025). ChatGPT-enhanced mobile instant messaging in online learning: Effects on student outcomes and perceptions. *Computers in Human Behavior, 168*, 108659.

*Jiang, Y. & Nakatani, K. (2025). Exploring implementations of GenAI in teaching IS Subjects and student perceptions. *Journal of Information Systems Education, 36*(2), 180-194.

*Lee, H. Y., Chen, P. H., Wang, W. S., Huang, Y. M., & Wu, T. T. (2024). Empowering ChatGPT with guidance mechanism in blended learning: Effect of self-regulated learning, higher-order thinking skills, and knowledge construction. *International Journal of Educational Technology in Higher Education*, *21*(1), 16.

*Li, H., Wang, Y., Luo, S., & Huang, C. (2024). The influence of GenAI on the effectiveness of argumentative writing in higher education: evidence from a quasi-experimental study in China. *Journal of Asian Public Policy, 18*(2), 405–430.

*Nasiar, N., Shah, M., Baker, R. S., & Pankiewicz, M. (2025). Scaffolding creativity or homogenizing outputs? GPT-based virtual TA's impact on student ideation in graduate education. *Social Innovations Journal*, *30*(2).

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *bmj*, 372.

Polya, G. (1945). *How to solve it: A new aspect of mathematical method.* Princeton university press.

*Rahman, G., Almutairi, E. A. A., Mudhsh, B. A., & Al-Yafaei, Y. (2025). Harnessing generative AI for collaborative creativity: A study of university students' engagement and innovation. International *Journal of Innovative Research and Scientific Studies*, *8*(3), 3284-3296.

*Shanshan, S., & Sen, G. (2024). Empowering learners with AI-generated content for programming learning and computational thinking: The lens of extended effective use theory. *Journal of Computer Assisted Learning, 40*(4), 1941–1958.

*Wang, C., Wang, B., & Xu, D. (2025). The role of ChatGPT and grammarly in promoting emotion regulation, psychological well-being, motivation, and academic writing in Chinese college students: A self-determination theory perspective. *Learning and Motivation*, *90*, 102131.

*Xu, X., Qiao, L., Cheng, N., Liu, H., & Zhao, W. (2025). Enhancing self-regulated learning and learning experience in generative AI environments: The critical role of metacognitive support. *British Journal of Educational Technology, 56*(5), 1842-1863.

International Society of the Learning Sciences

Technological Competency: A Conceptual Model for Understanding the World Around Us

Matthew Duvall, Chen Wang
mduv@upenn.edu, cwang0707@outlook.com
University of Pennsylvania – Graduate School of Education

Abstract: This conceptual paper describes a framework for fostering "technological competency" (TC), integrating ideas from multiple disciplines including literacies, computer science, and computer supported collaborative learning as a social practice, data literacy, AI literacy, and justice-centered computer science education. TC is the ability to (1) ask questions about technology that are relevant both personally and collectively; (2) identify, collect, and curate data to answer these questions; (3) use appropriate tools to analyze this data; and (4) report and re-evaluate this data in discussion with relevant communities. Most importantly, this should be collaborative work between learners of all kinds - researchers, practitioners, students, and other stakeholders. We conclude with practical guidelines for designers of learning experiences: stakeholder collaboration, diversity and inclusion, critical approaches, and effective communication.

Introduction

In the learning sciences, there are calls for improving data literacy learning, teaching future computer scientists to think more critically, and of course a growing call for incorporating AI literacy throughout the educational system. However, for practitioners in K-12 settings, it can be difficult to wade through all of this information to arrive at practical solutions for teaching these concepts to the real students sitting in their classrooms every day.

This conceptual paper describes "technological competency" (TC) - an understanding of how the technological systems and tools that impact our lives are shaped by the inherent social structures and human perspectives that underlay them. TC is a way of thinking about various literacies and learning approaches related to technology and its role in our lives. TC is the ability to (1) ask critical questions about technology and its impact that are relevant and important both personally and collectively; (2) identify, collect, and curate data to answer these questions in context; (3) use appropriate tools and techniques to analyze this data; and (4) communicate, evaluate, and refine the questions and answers in discussion with various relevant contexts..

By framing TC as both a conceptual model and a practical framework, we aim to support research-practice partnerships that connect disciplinary expertise with classroom needs. Aligned with the conference theme *Partnering with Purpose*, TC can serve as a shared foundation for cross-sector collaborations, enabling more intentional, equity-driven learning experiences that are rooted in both socio-technical understanding and everyday classroom practice.

Theoretical perspectives

Literacy is often viewed as a collection of discrete skills that can be acquired step-by-step (Hamilton, 2010; Hardy et al. 2022). In the early 1980's, literacy scholars including Shirley Brice Heath pointed out that observable literacy events, such as reading a book, are part of broader social interactions. The concept of literacy as a social practice (Hamilton, 2010; Perry, 2012) acknowledges that while literacy includes certain knowledge and skills, it is also defined by broader social and cultural implications. This also means that literacy can be viewed in different ways by different people at the same time, and that what it means to be literate may change over time as society and culture change.

A world powered by computerized systems results in a lot of data being generated and stored. As a result, Hardy et al. (2022) and Rubin (2022) called on education to broaden the view of data literacy from a conceptual data approach - teaching things like inference, variability, and visualization from specific content areas such as mathematics or biology - to also include a critical data literacy approach that positions students as not just consumers, but producers of data. This view empowers humans to not just calculate, but also critique, data practices and use their lived experiences to inform how/why they interact with these data.

The explosion of generative AI in 2022 has only added to an already growing call to action for schools and teachers to include AI literacy in the curriculum. For example, Casal-Otero et al. (2023) conducted a systematic literature review and described the need for better assessment of students' understanding of AI as well as a more intentional focus on the negative consequences of haphazard use of AI. While there is a growing sense

 © ISLS

of urgency about teaching AI literacy, there is still confusion and debate about how to define it, let alone how it should be taught to students from kindergarten to high school.

Related to these literacy concepts are ideas about restructuring how we teach computer science to our students. Vakil (2018) has called for a shift to justice-centered equity: to consider and critique the ethical uses of technology and computing, along with the rights and responsibilities of individuals and organizations within our society; to include opportunities for students to adopt disciplinary ideas in computer science while also creating environments that recognize and value the students' own social, civic, and political identities; and finally to connect computer science to political vision, acknowledging that computer science is important for the economic and political welfare of informed citizens.

While GenAI points to a new perspective on "learning with computers" (Salomon et al. 1991), there is a scramble to define AI literacy and best practices for design and use of such tools. In this landscape, there is a need for Principled Practical Knowledge (Bereiter 2014) to bridge the gap between theory and practice. TC is more than just an AI literacy framework. In fact, while TC is a concept that developed in response to questions about GenAI and how to teach young people about these tools, our goal is a more robust model that can describe ways to interact with computers and other humans in ways that are personally and communally productive. Considering concerns about technology over the last few decades (e.g computational thinking, CS for all, 21st century skills, the Metaverse, virtual reality, and so on), it seems relevant to develop a more general approach that can accommodate the current and future technological landscape, whatever that may be.

In the spirit of Wise and Schwarz (2017), we position TC within CSCL as a collaboratively developed and socially constructed practice that encompasses both humans and machines interacting in collaboration with each other. Our approach supports learners in engaging with real-world, complex aspects and questions about gen-AI and other technology through collaborative data practices and ethical reflection. Moreover, TC can guide the design of future CSCL environments that aim to foster critical engagement with technology and promote learning.

Our concept of TC is influenced indirectly by many more perspectives that space prohibits us from detailing here - researchers such as Papert, Wing, Kafai, and more. TC is not a replacement framework, but a meta-orchestrating construct that integrates inquiry, data practice, analysis, and community-based reflection across different technological contexts.

Discussion

Humans need TC to be more fully participating members in society. This broad idea builds from conceptions of literacy as a social practice, data literacy, AI literacy, CS education, and CSCL. Our current vision of TC includes four elements. (1) Ask critical questions: draws from traditions of critical literacy and justice-centered computer science education, which foreground issues of power, representation, and ethical responsibility in socio-technical systems. (2) Collect data: grounded in critical data literacy, emphasizing learners' roles as producers and interpreters of data rather than passive recipients. (3) Analyze data: builds on computational thinking and AI literacy perspective, which help understanding how technological systems operate and how analytic processes shape knowledge construction. (4) Communicate and discuss: reflects on CSCL and literacy frameworks, positioning learning as socially situated, dialogic, and collaboratively constructed. TC is not a binary "yes/no" skill, but rather one that can be applied in better or worse ways. TC is also a skill that is accessible, in some fashion, to any learner at any level.

Conclusion

This discussion of TC emphasizes the importance of helping learners think critically about the technology and tools we use every day. As technology continues to permeate education, work, and daily life, it is essential that learning experiences foster inquiry, reflection, and agency, and TC offers a lens for designing such experiences. If TC describes the competencies learners should develop, then designers of learning environments must intentionally structure opportunities for collaborative inquiry, data engagement, and critical reflection. From our initial conceptualization of this model, here are guidelines for designers of learning experiences/tools: *(1) Stakeholder collaboration:* Researchers, practitioners, students, and community members should work together to understand technology and its impact on their lives. *(2) Diversity and inclusion:* Asking and answering questions about technology should prioritize diverse perspectives by including a wide range of team members. Such diversity can lead to more inclusive and complete outputs. *(3) Critical approaches:* All affected stakeholders should have the opportunity to be producers, not just consumers, and should be encouraged to think critically about how and why to use technology for learning and creating. *(4) Effective communication:* The questions and answers should be communicated in ways that invite reflection and critique, with a view to establishing discourse and improving collective knowledge-building, in both academic and other settings.

 © ISLS

References

Bereiter, C. (2014). Principled practical knowledge: Not a bridge but a ladder. *Journal of the learning sciences, 23*(1), 4-17.

Casal-Otero, L., Catala, A., Fernández-Morante, C., Taboada, M., Cebreiro, B., & Barro, S. (2023). AI literacy in K-12: a systematic literature review. *International Journal of STEM Education, 10*(1), 29.

Hamilton, M. (2010). The social context of literacy. Teaching adult literacy: *Principles and practice,* 7-27.

Hardy, L., Dixon, C., & Hsi, S. (2022). From data collectors to data producers: Shifting students' relationship to data. In *Situating Data Science: Exploring How Relationships to Data Shape Learnin*g (pp. 104-126). Routledge.

Perry, K. H. (2012). What is literacy?--A critical overview of sociocultural perspectives. *Journal of Language and Literacy Education, 8*(1), 50-71.

Rubin, A. (2022). Learning to reason with data: How did we get here and what do we know?. In Situating Data Science: Exploring How Relationships to Data Shape Learning (pp. 154-164). Routledge.

Vakil, S. (2018). Ethics, identity, and political vision: Toward a justice-centered approach to equity in computer science education. *Harvard Educational Review, 88*(1), 26-52.

Salomon, G., Perkins, D. N., & Globerson, T. (1991). Partners in cognition: Extending human intelligence with intelligent technologies. *Educational researcher, 20*(3), 2-9.

Wise, A. F., & Schwarz, B. B. (2017). Visions of CSCL: Eight provocations for the future of the field. *International Journal of Computer-Supported Collaborative Learning, 12*(4), 423-46.

 © ISLS

International Society of the Learning Sciences

Idea Tracing Analysis (ITA): A Temporal-Structural Methodology for Modeling Collaborative Discourse and Contextualizing Multimodal Interactions

Xiaomeng Huang, Xavier Ochoa, Madhumitha Gopalakrishnan, Shiyu Zhang
xiaomeng.huang@nyu.edu, xavier.ochoa@nyu.edu, mg7584@nyu.edu, sz3561@nyu.edu
New York University

Abstract: Understanding learners 'behaviors in collaborative problem solving (CPS) is critical yet challenging due to its dynamic, multimodal nature. We introduce Idea Tracing Analysis (ITA), a method that models collaborative discourse as the evolution of interconnected ideas. ITA represents interactions as temporally aligned idea networks, capturing how ideas emerge, transform, and connect across participants. This approach provides a temporal and structural lens on collaboration, enabling analysis of group-level knowledge-building trajectories and the contextualization of multimodal interactions.

Introduction

One of the central goals in collaborative problem-solving (CPS) research is to understand how students act and interact—how they construct joint problem spaces or establish common ground through both verbal and nonverbal behaviors (Graesser et al., 2018; Stahl & Hakkarainen, 2021; Suthers et al., 2010; Wise et al., 2021). A deeper understanding of these interactional processes enables the design of more effective CPS activities and actionable supports (Jung et al., 2025). However, CPS processes are inherently *nested* and *situated* (Borge & Mercier, 2019). *Nested* refers to the interdependence between actions occurring at multiple levels, such as individual and group cognition (Barron, 2003; Stahl, 2005). *Situated* highlights that collaboration unfolds dynamically through moment-to-moment interactions in which common ground is continuously negotiated and updated (Cornelius & Herrenkohl, 2004; Graesser et al., 2018). At the same time, face-to-face collaboration is multimodal, involving speech, gesture, gaze, and interactions with artifacts. Multimodal learning analytics (MmLA) has enabled researchers to capture these channels and infer key collaboration constructs (Cukurova et al., 2018; Ochoa, 2022). However, interpreting multimodal behaviors without grounding them in an interactional context remains challenging.

To address this gap, we propose Idea Tracing Analysis (ITA), a methodology that models collaborative discourse through the temporal and structural evolution of ideas. This is built on CSCL perspectives that emphasize idea improvement as central to collaboration (Scardamalia & Bereiter, 2005). Prior work has examined ideas as units of analysis through uptake (Suthers et al., 2010) and idea-based metrics such as distribution and similarity (Chen et al., 2015; Zhu & Chen, 2023). ITA extends this work by tracing how ideas are introduced, elaborated, connected, and transformed over time, capturing their evolving relationships. By representing discourse as dynamic idea structures, ITA identifies collaboration contexts across multiple timescales and informs the interpretation of multimodal behaviors. In this way, ITA positions ideas as a bridge between interactional contexts and individual behaviors, providing a quantitative yet contextualized means of linking multimodal indicators with constructs, a connection often missing in current multimodal collaboration analytics (Huang & Ochoa, 2025).

Idea Tracing Analysis (ITA): An illustrated description

We introduce Idea Tracing Analysis (ITA) through an illustrative walkthrough. Figure 1 provides an overview of the process, which unfolds across four stages. We begin with *Stage 0*: Data Preparation, where speech is transcribed and multiple streams of multimodal data are synchronized into a unified dataset. In *Stage 1*, ideas are coded for their relationships and evolution stages, while multimodal behavioral features are extracted. *Stage 2* identifies collaboration contexts at the micro-, meso-, and macro-levels through idea metrics. Finally, *Stage 3* situates multimodal behaviors within these contexts to uncover behavioral patterns and construct learner profiles that reflect the nested and dynamic nature of collaboration.

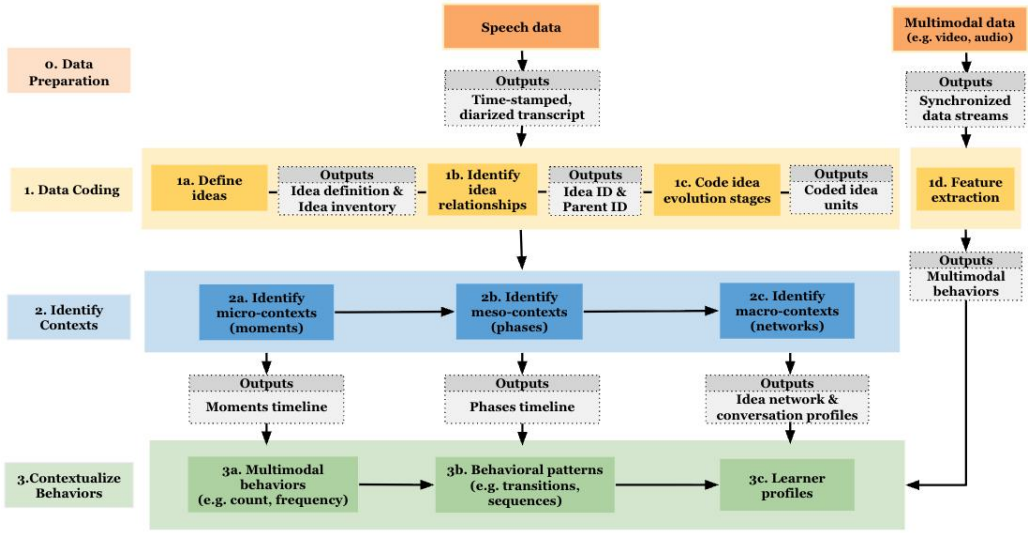
To illustrate ITA, we draw on data from a study of active listening in CPS, where 30 university students collaborated in groups of three to four on a 20–30 minute design challenge. Their interactions were captured through audio, video, and smart pen data. Figure 2 presents a macro-level example of idea evolution as a proposer–idea network, where nodes represent ideas and directed links capture how ideas build on one another. Each idea is linked to its proposer, enabling analysis of both idea development and the distribution of contributions across

 © ISLS

group members. Shaded regions indicate meso-level collaboration phases, allowing examination of how ideas evolve across temporal and structural dimensions.

Figure 1

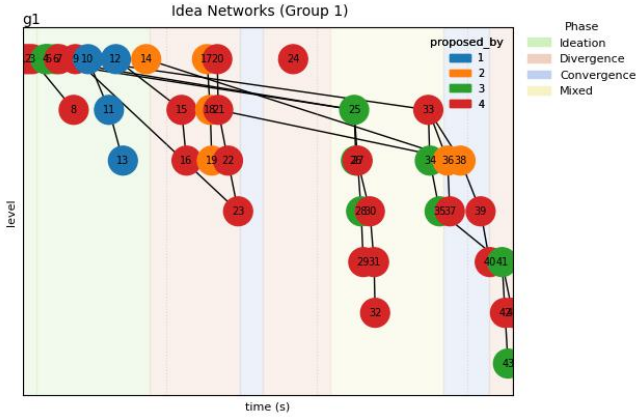
High-level Overview of Idea Tracing Analysis (ITA) Methodology



This analysis shows how collaboration unfolds over time by tracing how ideas are proposed, elaborated, connected, and taken up across participants. It reveals overall patterns of collaboration, such as how balanced, cohesive, or fragmented group interactions are. These identified contexts further support the interpretation of multimodal behaviors—for example, when a participant builds on others 'ideas, how this verbal action reshapes idea structures, and how such processes unfold over time.

Figure 2

The Temporal and Structural Evolution of Ideas (Example)



Conclusion

This paper introduces Idea Tracing Analysis (ITA), a methodology that traces idea evolution to identify collaboration contexts and situate multimodal behaviors within them. Building on prior work in collaborative discourse analysis and MmLA, ITA moves beyond frequency- and sequence-based approaches to reveal when and where behaviors such as gaze, gesture, and verbal actions become meaningful. ITA contributes to collaborative discourse analysis by treating ideas not only as indicators of collaborative processes but also as an analytic lens for identifying interactional contexts and interpreting multimodal behaviors beyond verbal discourse. It advances MmLA by situating behaviors within multi-level contexts and offering a systematic pipeline—from data preparation to context identification and behavior interpretation.

 © ISLS

References

Barron, B. (2003). When Smart Groups Fail. *Journal of the Learning Sciences*, *12*(3), 307–359. https://doi.org/10.1207/S15327809JLS1203_1

Borge, M., & Mercier, E. (2019). Towards a micro-ecological approach to CSCL. *International Journal of Computer-Supported Collaborative Learning*, *14*(2), 219–235. https://doi.org/10.1007/s11412-019-09301-6

Chen, B., Scardamalia, M., & Bereiter, C. (2015). Advancing knowledge-building discourse through judgments of promising ideas. *International Journal of Computer-Supported Collaborative Learning*, *10*(4), 345–366. https://doi.org/10.1007/s11412-015-9225-z

Cornelius, L. L., & Herrenkohl, L. R. (2004). Power in the Classroom: How the Classroom Environment Shapes Students 'Relationships With Each Other and With Concepts. *Cognition and Instruction*, *22*(4), 467–498. https://doi.org/10.1207/s1532690Xci2204_4

Cukurova, M., Luckin, R., Millán, E., & Mavrikis, M. (2018). The NISPI framework: Analysing collaborative problem-solving from students 'physical interactions. *Computers & Education*, *116*, 93–109. https://doi.org/10.1016/j.compedu.2017.08.007

Graesser, A. C., Fiore, S. M., Greiff, S., Andrews-Todd, J., Foltz, P. W., & Hesse, F. W. (2018). Advancing the Science of Collaborative Problem Solving. *Psychological Science in the Public Interest*, *19*(2), 59–92. https://doi.org/10.1177/1529100618808244

Huang, X., & Ochoa, X. (2025). Charting the Development of Collaboration Skills Through Collaborative Learning Analytics Systems. *Journal of Learning Analytics*, *12*(1), Article 1. https://doi.org/10.18608/jla.2025.8523

Jung, Y., Zhu, X., Oshima, J., Oshima, R., Chen, B., Moon, J., McNeill, L., Edmonds, C., Banihashem, K., Noroozi, O., Dey, I., Puntambekar, S., & Jackson, T. J. (2025). Towards Actionable Collaborative Discourse Analysis: Bridging Advanced Computational Analysis with Practical Implementation. *Proceedings of the 18th International Conference on Computer-Supported Collaborative Learning-CSCL 2025*, 472–480.

Ochoa, X. (2022). Chapter 6: Multimodal Learning Analytics—Rationale, Process, Examples, and Direction. In *The Handbook of Learning Analytics* (pp. 54–65). Soc. Learn. Analytics Res.

Scardamalia, M., & Bereiter, C. (2005). Knowledge Building: Theory, Pedagogy, and Technology. In R. K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (1st ed., pp. 97–116). Cambridge University Press. https://doi.org/10.1017/CBO9780511816833.008

Stahl, G. (2005). Group cognition in computer-assisted collaborative learning. *Journal of Computer Assisted Learning*, *21*(2), 79–90. https://doi.org/10.1111/j.1365-2729.2005.00115.x

Stahl, G., & Hakkarainen, K. (2021). Theories of CSCL. In U. Cress, C. Rosé, A. F. Wise, & J. Oshima (Eds.), *International Handbook of Computer-Supported Collaborative Learning* (pp. 23–43). Springer International Publishing. https://doi.org/10.1007/978-3-030-65291-3_2

Suthers, D. D., Dwyer, N., Medina, R., & Vatrapu, R. (2010). A framework for conceptualizing, representing, and analyzing distributed interaction. *International Journal of Computer-Supported Collaborative Learning*, *5*(1), 5–42. https://doi.org/10.1007/s11412-009-9081-9

Wise, A. F., Knight, S., & Shum, S. B. (2021). Collaborative Learning Analytics. In U. Cress, C. Rosé, A. F. Wise, & J. Oshima (Eds.), *International Handbook of Computer-Supported Collaborative Learning* (pp. 425–443). Springer International Publishing. https://doi.org/10.1007/978-3-030-65291-3_23 .

Zhu, X., & Chen, B. (2023). *Understanding Idea Creation in Collaborative Discourse through Networks: The Joint Attention-Interaction-Creation (AIC) Framework* (arXiv:2305.16262). arXiv. https://doi.org/10.48550/arXiv.2305.16262

International Society of the Learning Sciences

Do Anti-democratic Bots Dream of Digital Riots? Modeling Anti-democratic Decision-making and Collective Knowledge Construction Using Natural Language Processing

Renato Russo, TLT Lab, Columbia University (USA), rfr2126@tc.columbia.edu
Carolina Sotério, TLT Lab, Columbia University (USA), cs4465@tc.columbia.edu
Fabio Campos, TLT Lab, Columbia University (USA), fabio.campos@tltlab.org
Paulo Blikstein, TLT Lab, Columbia University (USA), Instituto Tecnológico de Monterrey (Mexico), paulob@tc.columbia.edu
Gustavo Oliveira, USP (Brazil), gustavo.oliveira03@usp.br

Abstract: We introduce a tool for visualizing reasoning trajectories in online interactions. We leverage natural language processing to corpora from Brazilian communities involved in anti-democratic speech to compare bots' and people's trajectories around "discourse crossroads" where epistemic stances towards democratic norms shift. We hope to inform tools and intervention design that guide epistemic trajectories toward more productive engagement, highlighting the role of civic education and digital literacy in fostering critical dialogue.

Introduction and literature review

Anti-democratic groups have promoted their agendas through practices that go beyond *simple* repetition of non-facts: they employ practices resembling *agential knowledge construction* and community-based epistemologies, but distorted (de Nadal, 2024; Lewis et al., 2021; Russo & Blikstein, 2022, 2025). The rise of QAnon, for example, reflects collective sense-making which scholars link with "communities of practice" (Ylä-Anttila, 2018; Lave & Wenger, 1991) and informal miseducation (Greenhalgh et al., 2021). In other words, a supposed rationality is used to promote harmful agendas online through sense-making that feels engaging and agential.

Our methodological contribution addresses a challenging issue: misinformation thrives when reasoning drifts away from reciprocity and contestability. That raises an *epistemic* problem, beyond a merely *factual* one. We introduce a tool for modeling and visualizing reasoning trajectories in online interactions. The tool focuses on *discursive tipping points*—junctures where epistemic stances shift from compliant or borderline to non-compliant with democratic norms. We frame misinformation as a *process of epistemic change* and ask: **how might we visualize the tipping points where epistemic stances depart from democratic principles?** Our replicable design (Figure 1) compares how bots and humans fare in a decision tree intentionally designed to present "ideological crossroads" that lead either to anti-democratic choices or to choices compliant with established frameworks for democratic policymaking in Brazil (our home country). We contribute to the study of the pedagogy of misinformation and the fragile reasoning processes that sustain democratic life online.

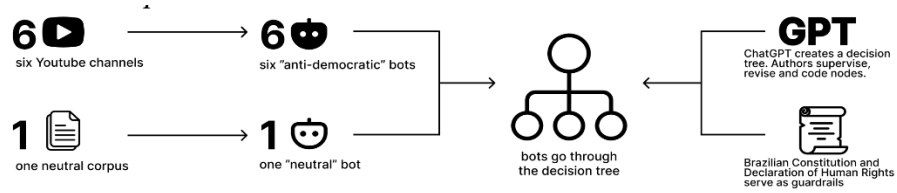
We build on previous research (Russo et al., 2024) to extend research analyzing large-scale online discourse to trace knowledge construction and epistemic change (Sherin & Chavey, 2025) and draw connections with two other strands of literature: public pedagogy scholarship demonstrating how anti-democratic discourse spreads online as forms of collective knowledge construction; (studies of social media as learning spaces.

Design rationale

We created a decision tree to be run through by seven bots (six anti-democratic ones and one "neutral"). The decision tree had several turning points where discourse "drifted" from compliant with democratic principles to non-compliant. Additionally, with the intent to simulate an analysis that juxtaposes robots' performance in the task with humans', we ran a preliminary study with a limited sample of human respondents. Brazilian participants from Prolific (N=54) performed a task similar to that performed by bots: given a root node, choose a response; from that response, choose 1 out of 3 options. Finally, participants indicated their political stance using a slider where 1 meant *Left* and 10 meant *Right*. Of special interest to the analysis of human respondents were the leaf nodes they chose and the frequency of each label (compliant, dubious, non-compliant).

Figure 1
A Summary of the Tree Design and Test Stages, Before Comparison of Results with that of Human Respondents

 © ISLS



Preliminary findings

We identified that *all "anti-democratic bots"* made the same three choices throughout the tree: After the parent node, all went to node #2 ("You can't trust anyone… media, politicians, everybody is feathering their own nest"), then node #7 ("Go check 2, 3 sources… There is always a different version spreading. Good information comes from comparison #SuspectAndMultiply), and then node #24 ("Comparing everything can be helpful, but it brings about confusion ⚖️"). The order of labels was, thus, *dubious* > *compliant* > *dubious*. In contrast, the *neutral bot followed only compliant nodes*: #3 ("Does traditional journalism only serve the boss? Let's make independent media stronger, then 📣"); #11 ("Media regulation could be the solution… but it depends: if there is no transparency, it may end up in censorship"); #34 ("Saw a news story? Let's take a look at another source. 🔎"). Most humans finished at a compliant node, and only 2 participants finished their trajectories at a dubious node (#24), one participant at each extreme of the political spectrum. Also, the majority chose *compliant nodes only*. No comparison is possible across political ideologies.

Intriguingly, about 1/4 of humans (mostly self-identified centrists) made the same first choice as the anti-democratic bots (node #2, "You can't trust anyone… media, politicians, everybody is feathering their own nest"). What made that node attractive to those bots and to all those humans? If—as we hypothesize—the bots truly represent stances collectively taken by the users whose voices make up the bots' decision-making, then, possibly, (i) node #2 might be acting as a *gateway premise* that leads viewers to statements that diverge from compliance with our frameworks; and/or (ii) the "collective intelligence" of YouTube comments might have already incorporated that belief into its "sense-making" supported by a blend of black-box algorithms and visible human user input. To extend our metaphor, ***some bots in the wild might already be dreaming of digital riots***—like the ones their human counterparts put in place in Brasilia in January 2023. It is also intriguing that crucial node's content evokes the erosion of trust in institutions, an emerging phenomenon associated with groups involved in the Capitol riots (e.g., Marwick & Partin, 2022).

Brief discussion and conclusion

We plan to extend the decision-tree framework by introducing finer-grained distinctions between nodes. We also aim to test the same decision-tree framework with larger human samples, representing a range of political alignments. By comparing human participants' and bots' behavior, we can examine whether our hypothesized tipping points hold across contexts and perspectives. This will provide critical validation for the model and highlight potential variations between automated and human reasoning. Together, these extensions strengthen the methodological contribution of our work by demonstrating the tool's capacity to trace reasoning pathways with greater precision. By operationalizing discursive tipping points and testing them with both bots and human participants, we aim to provide researchers and educators with a replicable resource to make visible the fragile mechanisms by which democratic reasoning can be reinforced—or disrupted—in online environments.

Our design allows us to trace trajectories through successive discourse junctures, foregrounding pathways along which reasoning develops. This hybrid framework creates new opportunities to analyze agency and critical engagement as learners navigate contested epistemic terrains. Results also raise design implications. Nudges, prompts, or collaborative tasks could be used to help learners sustain productive engagement even when exposed to manipulative discourses. *Importantly, these findings remind us that simply providing corrective facts or additional evidence is insufficient.* What matters is how we scaffold reasoning, how standards of justification are upheld, and how learners position themselves for accountability within pluralistic dialogue.

References

de Nadal, L. (2024). From Denial to the Culture Wars: A Study of Climate Misinformation on YouTube. *Environmental Communication*, 1–18. https://doi.org/10.1080/17524032.2024.2363861

Greenhalgh, S., Krutka, D., & Oltmann, S. (2021). Gab, Parler, and (Mis)educational Technologies: Reconsidering Informal Learning on Social Media Platforms. *Journal of Applied Instructional Design*,

10(3). https://doi.org/10.51869/103/sgdkso.

Lave, J., & Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge University Press.

Lewis, R., Marwick, A. E., & Partin, W. C. (2021). "We Dissect Stupidity and Respond to It": Response Videos and Networked Harassment on YouTube. *American Behavioral Scientist*, *65*(5), 735–756. https://doi.org/10.1177/0002764221989781

Russo, R., & Blikstein, P. (2022). *Beyond Belief: Disinformation Strategies and Collective Knowledge Building Led by Jair Bolsonaro in Brazil*. https://repository.isls.org//handle/1/8540

Russo, R., & Blikstein, P. (2025). Towards a Learning Sciences-Informed Framework for Investigating "Twisted" Knowledge Construction on Social Media. *Proceedings of the 19th International Conference of the Learning Sciences - ICLS 2025*. 19th International Conference of the Learning Sciences - ICLS 2025.

Russo, R., Blikstein, P., & Literat, I. (2024). Twisted knowledge construction on X/Twitter: An analysis of constructivist sensemaking on social media leading to political radicalization. *Information and Learning Sciences*, *125*(9), 693–719. https://doi.org/10.1108/ILS-12-2023-0210

Sherin, B., & Chavey, L. (2025). Learning Trajectories in Large-Scale Communities. *Proceedings of the 19th International Conference of the Learning Sciences - ICLS 2025*, 1500–1504. https://doi.org/10.22318/icls2025.721043.

Ylä-Anttila, T. (2018). Populist knowledge: 'Post-truth' repertoires of contesting epistemic authorities. *European Journal of Cultural and Political Sociology*, *5*(4), 356–388. https://doi.org/10.1080/23254823.2017.1414620

 © ISLS

International Society of the Learning Sciences

Exploring Collaborative Teacher-AI Interaction for Rehearsing Responsive Teaching in Elementary Mathematics Education

Shuman Wang, Gordon Yeung, Miroslav Suzara, Jennifer Osuna, Charis Charitsis, John Mitchell
shuman@stanford.edu, wlyeung@stanford.edu, msuzara@stanford.edu, jmlo@stanford.edu,
charis@stanford.edu, John.Mitchell@stanford.edu
Stanford University

Abstract: This proposal introduces *TeacherCollab*, an AI-supported simulated professional learning community designed to cultivate early-career teachers' (ECTs) responsive teaching in elementary mathematics. Rather than simulating students, *TeacherCollab* positions AI as collaborative teacher colleagues who model diverse pedagogical lenses while reasoning around authentic students' work. Grounded in approximations of practice and computer-supported collaborative learning (CSCL) theory, *TeacherCollab* creates a multi-agent dialogue environment where ECTs interact with AI-generated personas that surface students' relational dynamics, mathematical strategies, and areas of struggle. A pilot study with 16 ECTs showed that *TeacherCollab* fostered engagement, reflection, and awareness of multiple perspectives on student learning.

Introduction

Responsive teaching, which requires teachers to notice, interpret, and respond to the substance of students' ideas, is a high-leverage instructional practice that early-career teachers (ECTs) find difficult to master (Sherin et al., 2011; Dyer & Sherin, 2016). To support ECTs, teacher educators have developed *approximations of practice*, simplified environments for rehearsing complex teaching moves (Grossman et al., 2009). More recently, large language models (LLMs) have been deployed to simulate student dialogue to support ECTs in eliciting and responding to students' thinking (e.g., Lee & Yeo, 2022; Dai et al., 2024; Lee et al., 2025). Early evaluations highlight clear affordances, including iterative and scalable rehearsal, more open-ended turns, and a stronger presence. However, they also surface limits that matter for transfer in actual teaching. For example, AI-simulated students can produce generic or implausible reasoning, struggle to offer productive beginnings for teacher uptake, and exhibit brittleness or repetition, which can nudge ECTs toward low-leverage moves (Zhang, 2024; Dai et al., 2024). These findings suggest reframing AI's role: rather than perfecting a single "AI student", simulations should be grounded in authentic student artifacts and layered with in-the-moment scaffolds that surface multiple interpretive lenses. In view of this, we introduce *TeacherCollab*, a teacher-facing simulation platform that moves from using AI to simulate students to positioning AI as collaborative teacher colleagues in a multi-agent, dialogic environment grounded in authentic student work. We introduce the *TeacherCollab* platform and explore the potential of an AI-stimulated platform to scaffold ECTs' engagement in the professional practice of making sense of students' mathematical thinking.

The TeacherCollab platform

TeacherCollab simulates a teacher professional learning environment (PLC) that invites ECTs to practice noticing and making sense of student mathematical thinking, and to do so in dialogue with AI-generated simulated teacher colleagues who collaborate with one another and with the novice teacher around the same artifact (Klieger et al., 2024). ECTs are invited to participate in teacher-AI interactions about noticing and making sense of students' work and dialogue. Each AI agent persona adopts a distinct pedagogical stance, which helps ECTs construct a rich, humanizing read of students' work and dialogue. In particular, one agent examines social and relational aspects of student collaboration, particularly talk-move patterns, such as positioning and revoicing (Barron, 2003; Chapin et al., 2009). A second attends to students' learning strategies, identifying and naming students' mathematical thinking, positioning every learner as a capable thinker (Carpenter et al., 1999; Jacobs et al., 2010). A third focuses on identifying students' areas of struggle involving partial understandings and confusion, and proposes respectful follow-ups that preserve student ownership while nudging thinking forward (Kapur, 2014; Warshauer, 2015). These lenses invite ECTs to develop a multidimensional, asset-oriented perspective they can apply to artifacts of student thinking, enabling responses that both advance mathematical ideas and humanize the learning process. Each session centers on a short lesson snapshot, a focused slice of classroom activity paired with the relevant students' work or dialogue, which anchors a brief sequence of guided dialogue. ECTs first engage one-at-a-time with an AI agent representing the respective interpretive lens (e.g., interactional dynamics, strategies, areas of struggle) (see Figure 1, left). The conversation then expands into a roundtable discussion, in

 © ISLS

which multiple agents speak together with the user. Contrasting interpretations are made explicit, and the user articulates a justified next instructional step that reconciles or prioritizes the perspectives (see Figure 1, right). The *TeacherCollab* interface is organized as a two-pane workspace (see Figure 2 and 3).

Figure 1
An Illustration of the Interaction Pattern of TeacherCollab

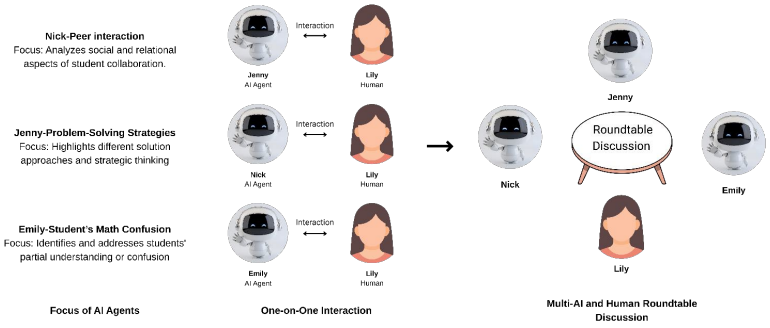


Figure 2
Tool Interface in Single-agent Mode

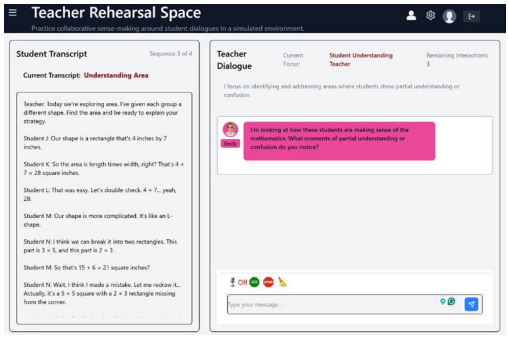
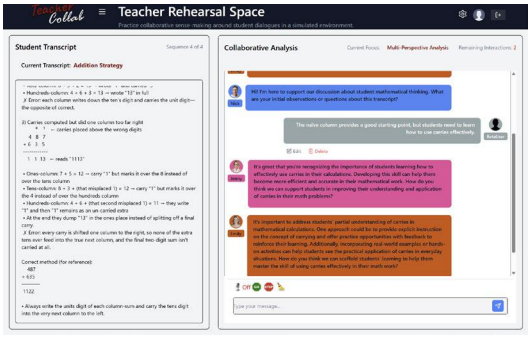


Figure 3
Tool Interface in Multi-agent Mode



Pilot study and discussion

We piloted *TeacherCollab* with participants (n=16). We collected session records, including lens choices, dialogue transcripts, and timestamps per snapshot, and collected post-case reflections from participants, where participants shared positive impressions of this tool. They appreciated the clear and accessible user interface and found the experience engaging. Participants noted that *TeacherCollab* prompted them to reflect on their own thinking and to consider diverse perspectives on student learning. 67% of participants reported feeling engaged throughout the process. In terms of learning potential, several participants highlighted that the tool enhanced their understanding of how children think mathematically and how social dynamics influence students' interactions. For improvement, participants recommended incorporating richer conversational variety as well as integrating multimodal student work, for example, using videos instead of only transcripts. *TeacherCollab* reframes AI's role from a simulator to a collaborative partner that scaffolds how teachers reason together around authentic student artifacts. The multi-agent design externalizes different interpretive lenses, creating productive tensions that prompt teachers to justify their pedagogical decisions. This aligns with CSCL's view of learning as knowledge building across multiple voices (Scardamalia & Bereiter, 2010). By focusing on augmenting professional vision rather than automating it, *TeacherCollab* offers a scalable model for developing the complex practice of responsive teaching. While the current study focused on ECTs because they are at a formative stage in developing professional noticing skills and stand to benefit most from structured scaffolding around interpreting students' thinking, the multi-agent design is not inherently limited to this population. Experienced teachers may engage differently with the tool (e.g., drawing on richer classroom repertoires to challenge or extend the AI agents' interpretations). Future research will examine how teachers at varying career stages interact with and benefit from the platform, incorporate multimodal artifacts (e.g., video clips), and examine how longitudinal engagement with the platform shapes teachers' professional vision and instructional practices over time.

 © ISLS

References

Barron, B. (2003). When smart groups fail. *The Journal of the Learning Sciences*, *12*(3), 307–359. https://doi.org/10.1207/S15327809JLS1203_1

Carpenter, T. P., Fennema, E., Franke, M. L., Levi, L., & Empson, S. B. (1999). *Children's mathematics: Cognitively guided instruction*. Heinemann.

Chapin, S. H., O'Connor, M. C., & Anderson, N. C. (2009). *Classroom discussions: Using math talk to help students learn, grades K-6*. Math Solutions.

Dai, C. P., Ke, F., Zhang, N., Barrett, A., West, L., Bhowmik, S., Southerland, S. A., & Yuan, X. (2024). Designing conversational agents to support student teacher learning in virtual reality simulation: A case study. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems* (pp. 1–8). Association for Computing Machinery. https://doi.org/10.1145/3613905.3637145

Dyer, E. B., & Sherin, M. G. (2016). Instructional reasoning about interpretations of student thinking that supports responsive teaching in secondary mathematics. *ZDM–Mathematics Education*, *48*(1), 69–82. https://doi.org/10.1007/s11858-015-0740-1

Grossman, P., Hammerness, K., & McDonald, M. (2009). Redefining teaching, re-imagining teacher education. *Teachers and Teaching: Theory and Practice*, *15*(2), 273–289. https://doi.org/10.1080/13540600902875340

Jacobs, V. R., Lamb, L. L. C., & Philipp, R. A. (2010). Professional noticing of children's mathematical thinking. *Journal for Research in Mathematics Education*, 41(2), 169–202. https://doi.org/10.5951/jresematheduc.41.2.0169

Kapur, M. (2014). Productive failure in learning math. *Cognitive Science*, *38*(5), 1008–1022. https://doi.org/10.1111/cogs.12107

Klieger, B., Charitsis, C., Suzara, M., Wang, S., Haber, N., & Mitchell, J. C. (2024). *ChatCollab: Exploring collaboration between humans and AI agents in software teams*. arXiv. https://doi.org/10.48550/arXiv.2412.01992

Lee, D., Son, T., & Yeo, S. (2025). Impacts of interacting with an AI chatbot on preservice teachers' responsive teaching skills in math education. *Journal of Computer Assisted Learning*, *41*(1), 1–19. https://doi.org/10.1111/jcal.13091

Lee, D., & Yeo, S. (2022). Developing an AI-based chatbot for practicing responsive teaching in mathematics. *Computers & Education*, *191*, 104646. https://doi.org/10.1016/j.compedu.2022.104646

Scardamalia, M., & Bereiter, C. (2010). A brief history of knowledge building. *Canadian Journal of Learning and Technology*, *36*(1). https://doi.org/10.21432/T2859M

Sherin, M. G., Jacobs, V. R., & Philipp, R. A. (Eds.). (2011). *Mathematics teacher noticing: Seeing through teachers' eyes*. Routledge. https://doi.org/10.4324/9780203832714

Warshauer, H. K. (2015). Productive struggle in middle school mathematics classrooms. *Journal of Mathematics Teacher Education*, *18*(4), 375–400. https://doi.org/10.1007/s10857-014-9286-3

Zhang, N., Ke, F., Dai, C.-P., Southerland, S. A., & Yuan, X. (2024). Seeking to support preservice teachers' responsive teaching: Leveraging artificial intelligence-supported virtual simulation. *British Journal of Educational Technology*, *56*(3), 1148–1169. https://doi.org/10.1111/bjet.13522

International Society of the Learning Sciences

The Impact of Conversational Agents on Children's and Parents' Engagement During Shared Reading

Dorentina Dedushaj, University of North Carolina at Chapel Hill, dedushaj@unc.edu
Kathryn Leech, University of North Carolina at Chapel Hill, leechk@unc.edu
Isabella Ruel, University of North Carolina at Chapel Hill, iruel@unc.edu
Grace Lin, Massachusetts Institute of Technology, gcl@mit.edu

Abstract: Engagement is an important aspect of parent-child shared reading interactions. However, little is known about the nature of engagement patterns while parent-child dyads engage with AI-powered CAs in at-home reading contexts. In this study, we analyze diverse observable patterns of engagement during shared reading. We examine the overall impact of an artificially intelligent conversational agent (CA) that augments parent-child reading interactions by observing and analyzing engagement patterns between parents and children as they read together.

Introduction

Interactive shared book reading promotes preschool-aged children's early literacy and future reading comprehension (Grøver et al., 2019; Mol & Bus, 2011). Conversational agents (CAs), a type of AI-powered digital technology that supports shared reading through real-time prompts, have the potential to enrich parent-child interactions around print books (Leech et al., 2025; Smeets & Bus, 2012). For effective shared book reading, both the parent and child need to be engaged (Lingwood et al., 2023; Oertel et al., 2021). Key indicators of children's engagement during reading are reflected in their emotional tone, nonverbal behaviors (e.g., pointing to the book or visually attending to it), and verbal attention (e.g., questions, answers, and comments), and their involvement and active participation (Fredricks et al., 2004; Strouse & Ganea, 2017). Engagement functions as a central mechanism in shared reading, shaping the quality of parent-child interaction and supporting learning and language development (Hirsh-Pasek et al., 2015; Strouse & Ganea, 2017). While prior studies have explored engagement during shared reading with CAs in lab settings, showing increased child participation and interaction (Xu & Warschauer, 2020; Xu et al., 2021), there is limited understanding of how AI-powered CAs impact at-home reading interventions and engagement patterns over time.

This study examines how AI-powered conversational agents shape behavioral engagement patterns in parent–child dyads during shared reading in at-home contexts over a four- to five-week period. The intervention group engaged in shared reading with a CA, whereas the control group engaged in traditional shared reading without the CA. This study investigates (RQ1) whether children and parents in the intervention group exhibit greater increases in engagement from Session 1 to Session 2 compared to those in the control group, and (RQ2) whether children's and parents' engagement differs between traditional shared reading and CA-assisted shared reading within the intervention group during Session 1.

Methods

The study included forty children (20 female, 20 male) between the ages of four and five years old who participated with a parent (38 female, 2 male). Children were predominantly White (n=24), with smaller representations of African American (n=7), Asian (n=4), Hispanic (n=4), and mixed race (n=1). Eligibility criteria included having a child in the age range (4;0 to 5;11), an annual household income of less than $75,000 in the previous year, speaking English at least 75% of the time at home, and having access to an Android smartphone and a reliable home internet. The average annual household income for participating families was $51,167 USD (SD = $17,698), with a median of $45,000 USD. Income was included as a proxy for socioeconomic status, which has been shown to influence variability in parent–child interaction and conversation.

Procedure

Participants from both conditions completed two 45-minute remote sessions over a four- to five-week period, which included at-home readings between sessions. Participants were randomly assigned to two conditions: the intervention group (n = 20) and the control group (n = 20). In Session 1, dyads in both groups first engaged in a traditional shared reading of *The Little Red Fort* (Brenda Maier) without the CA. The intervention group then completed a second book reading, *How to Find an Elephant* (Kate Banks), with the CA. The control group did

 © ISLS

not engage in CA-assisted reading in Session 1. In Session 2, conducted approximately five weeks after Session 1, dyads in both groups completed both a traditional reading and a CA-assisted reading interaction.

The app consisted of an AI voice-activated reading companion that listens to parent-child reading interactions and provides prompts to support shared reading. The system used voice-recognition technology to detect predefined marker words embedded in the text, defined as specific target words or phrases from the book that, when spoken aloud during reading, triggered the app's virtual character to produce a chime and provide a prompt tailored to the book's content and context (e.g., '*What would you need to find an elephant?'*). Both CA-assisted and traditional reading interactions were video-recorded for later transcription and coding.

Measure

To analyze children's and parents' behaviors, the shared book reading videos were trimmed into 30-second intervals and coded using ELAN (Version 7.0; Max Planck Institute for Psycholinguistics, 2025). Two research assistants coded each 30-second interval for four dimensions of child engagement (active participation, availability for reading, affective state, child involvement), as well as parent involvement, which captured the parent's engagement with and scaffolding of the child during shared reading. Scores ranged from 0 to 4, where 0 indicates the lowest level of engagement and 4 the highest. Each recording captures the parent, the child, and the book. If the child or parent went off camera, the research assistants coded the interval as 0. The coding scheme used in this study was adapted from Strouse and Ganea (2017), incorporating elements from the Leuven Involvement Scale for Young Children (Laevers, 1994). Inter-rater reliability was established on 50% of the sample between the two raters with substantial agreement, $\kappa = .74$ (Active Participation), $\kappa = .71$ (Affective State), $\kappa = .72$ (Child Involvement), $\kappa = .73$ (Parent Involvement), $\kappa = .69$ (Availability for Reading), all $p < .01$. The remaining data were then coded independently by the two raters.

Preliminary results

Session 1 coding is complete, and Session 2 coding is currently in progress. Preliminary descriptive analyses for Session 1 indicate that children's active participation was higher in the CA-assisted reading (M = 1.52) than in traditional reading (M = 0.95). Affective state showed a similar pattern (M = 2.35 vs. 1.96), and child involvement (M = 1.94 vs. 1.29) and parent involvement (M = 1.83 vs. 1.07) were also higher in the CA condition. In contrast, visual attention was higher during traditional reading (M = 2.90) than during CA-assisted reading (M = 2.17). Paired-samples *t*-tests indicated that affective state, active participation, child involvement, and parent involvement were significantly higher in the CA condition, whereas visual attention was higher in the traditional condition (*p*s < .05). Upon completion of Session 2 coding, we will conduct within-subject and between-condition analyses to examine engagement patterns across conditions and changes from Session 1 to Session 2.

Discussion

Preliminary findings indicate that CA-assisted shared reading was associated with higher children's active participation, affective engagement, and involvement, as well as greater parent involvement, compared to traditional shared reading. In contrast, children demonstrated higher visual attention during traditional reading. Prior work suggests that CAs can scaffold interactive reading exchanges, and earlier work with this CA found higher parent-child conversation during CA-supported reading (Xu et al., 2021; Xu & Warschauer, 2020; Leech et al., 2025). A strength of this design is the within-subjects comparison, which provides stronger evidence for the findings, as the same children showed different patterns of engagement depending on the presence of the CA.

Implications

We identified multiple areas for future research. While our preliminary findings suggest a positive relation between CA-assisted reading and engagement, longitudinal studies are needed to assess long-term effects. As shared reading is tied to later academic success (Grøver et al., 2019; Mol & Bus, 2011), this study suggests opportunities for CAs to be extended to formal learning settings, such as pre-school settings, to further assess engagement patterns, particularly in shared reading between teachers and students. Careful, ethical, and equitable considerations need to be taken when designing these tools to avoid bias, discrimination, or privacy violations. This study provides evidence that affordable, accessible, and effective tools that align with evidence-based literacy practices can be implemented across different learning settings, and these tools can benefit and reach children from diverse backgrounds. These data can be used to inform educational interventions and focus on maintaining social interaction between adults and children, all while supporting children's language development, emergent literacy skills, and later academic success. The findings from this study provide insight into the design of innovative and equitable CAs that enhance children's academic success.

References

ELAN (Version 7.0) [Computer software]. (2025). Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Retrieved from https://archive.mpi.nl/tla/elan

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. Review of Educational Research, 74(1), 59–109.

Grøver, V., Størksen, I., & Lunde, T. (2019). Shared book reading and children's language development in early childhood education and care: A study of Norwegian preschools. Early Childhood Research Quarterly, 48, 235–247. https://doi.org/10.1016/j.ecresq.2019.03.001

Hirsh-Pasek, K., Zosh, J. M., Golinkoff, R. M., Gray, J. H., Robb, M. B., & Kaufman, J. (2015). Putting education in "educational" apps: Lessons from the science of learning. Psychological Science in the Public Interest, 16(1), 3–34. https://doi.org/10.1177/1529100615569721

Laevers, F. (1994). The Leuven Involvement Scale for Young Children: Manual and video. Leuven, Belgium: Centre for Experiential Education.

Leech, K. A., Lin, G., Pedonti, S., Schoenfeld, I., & Uz Bilgin, C. (2025). Evaluating the usability of a conversational agent to enhance parent–child shared reading interactions. International Journal of Child-Computer Interaction, 43, 100720. https://doi.org/10.1016/j.ijcci.2024.100720

Lingwood, J., Lampropoulou, S., De Bezenac, C., Billington, J., & Rowland, C. F. (2023). Children's engagement and caregivers' use of language-boosting strategies during shared book reading: A mixed methods approach. Journal of Child Language, 50(6), 1436–1458. https://doi.org/10.1017/S0305000922000290

Mol, S. E., & Bus, A. G. (2011). To read or not to read: A meta-analysis of print exposure from infancy to early adulthood. Psychological Bulletin, 137(2), 267–296. https://doi.org/10.1037/a0021890

Oertel, C., Castellano, G., Cohn, J. F., & Campbell, N. (2021). The role of engagement in human–human and human–robot interaction: A survey. Frontiers in Robotics and AI, 8, 747–780. https://doi.org/10.3389/frobt.2021.747780

Smeets, D. J. H., & Bus, A. G. (2012). Interactive electronic storybooks for kindergartners to promote vocabulary growth. Journal of Experimental Child Psychology, 112(1), 36–55. https://doi.org/10.1016/j.jecp.2011.12.003

Strouse, G. A., & Ganea, P. A. (2017). Parent–toddler behavior and language differ when reading electronic and print picture books. Frontiers in Psychology, 8, 677. https://doi.org/10.3389/fpsyg.2017.00677

Xu, Y., & Warschauer, M. (2020). Exploring young children's engagement in joint reading with a conversational agent. Proceedings of the Interaction Design and Children Conference, 216–228. https://doi.org/10.1145/3392063.3394417

Xu, Y., Aubele, J., Vigil, V., Bustamante, A. S., Kim, Y. S., & Warschauer, M. (2021). Dialogue with a conversational agent promotes children's story comprehension via enhancing engagement. Child Development, 93(2), e149–e167. https://doi.org/10.1111/cdev.13708

Acknowledgments

The authors would like to thank the research assistants for their support, the advisor for guidance, and the children and parents for their time and participation.

International Society of the Learning Sciences

Novel Figurative Language to Identify Promising Ideas in Knowledge Building Scientific Discourse

Wakana Tsuji, Marlene Scardamalia, Ahmad Khanlari
wakana.gen@mail.utoronto.ca, marlene.scardamalia@utoronto.ca, a.khanlari@mail.utoronto.ca
University of Toronto

Abstract: This research explored grade 4 students' novel figurative language to identify promising ideas in knowledge building community discourse. A mixed-method approach—semantic analysis using Word2Vec and qualitative coding—was adopted for the analysis. Results showed figurative language use involved *conceptual reframing* across science and everyday life domains and *embodied perspective-taking*. Our next step is to incorporate similarity to resource material and investigate how ideas containing figurative language supports the community discourse.

Figurative language and creativity in scientific discourse

Figurative language, such as metaphors, play a crucial role in understanding and communicating science concepts (Cameron, 2002), through conceptual mapping of a familiar source domain onto a more complex, abstract target domain (Lakoff & Johnson, 1980). Creating associative connections between remote concepts has been characterized as a capacity for creative thinking (Mednick, 1962). Computational and neuroimaging studies confirmed a strong correlation between semantic distance in associative word pairs and creativity (Beaty & Johnson, 2021; Liu et al., 2020). Further, a recent study reported that semantic distance between mapped concepts aligns with creativity ratings in metaphors (Yang et al., 2025). Existing evidence is largely derived from elicited data; semantic distance may serve as a quantifiable indicator of creative metaphor and ideation in naturally occurring discourse.

Studies report low levels of metaphor production in the writing of elementary school students (Gardner et al., 1975; Thomas et al., 2025), particularly novel metaphors (Pollio & Pollio, 1974). Novelty (e.g., originality, uniqueness) and appropriateness (e.g., usefulness, value) are two qualities of creativity repeatedly discussed in literature (e.g., Runco & Jaeger, 2012). In knowledge building communities, theory development requires drawing on one's experience to map the familiar with complex, abstract, theoretical, and distant concepts and create new explanations and accounts. From a socio-cognitive perspective, students are dealing with difficult-to-understand concepts in a community context, trying to communicate to peers. In such contexts, students this age and younger have demonstrated the capacity to exceed curriculum expectations, engaging in metadiscourse and demonstrating significant vocabulary development (Khanlari et al., 2019; Resendes et al., 2015). These advances require sociocultural creativity (Glăveanu, 2012): novel conceptual and linguistic affordances are created through social judgement and acceptance (Amabile, 1988) of their value and usefulness to the group. This process is inherently cyclical, as new language develops through divergent alternatives and convergence through adaptation to community norms (Piccardo, 2017).

Student discourse is captured in Knowledge Forum technology built to support Knowledge Building. Theory-building epistemic markers (e.g., "my theory," "a better theory," "new information," "this theory cannot explain," "I need to understand," "putting our knowledge together") may foster novel figurative language despite its rarity in grade 4. We conducted the current pilot data analysis to explore the potential of figurative language serving as means to identify promising ideas, asking the following research question: C*an novel figurative language serve as means to identify promising ideas with epistemic value to the community in knowledge building community discourse?*

Methods

The data source was 132 Knowledge Forum archived notes on light and optics, produced by 23 grade 4 students over three months. We adopted a mixed-methods approach for data analysis: semantic analysis and qualitative coding. First, a semantic analysis was conducted using Word2Vec, a neural network–based natural language processing technique that represents words as continuous vectors in a high-dimensional space (Mikolov et al., 2013). This enabled the calculation of cosine similarity between all noun combinations within each sentence unit in notes with theory-building epistemic markers to measure their semantic relatedness on nouns, with values ranging from 0 (semantically close) to 1 (semantically distant). Pairs containing overly general words (e.g., thing, stuff), compound nouns, noun phrases, and pairs without syntactic relation were excluded. In addition, we excluded sentences that appeared to be repetition of figurative language used in source material. Since resource

 © ISLS

material was not available, we could not check semantic relatedness to it, so this procedure may not fully rule out its influence. This limitation will be addressed in future research. Procedures resulted in a total of 320 pairs included for qualitative analysis. Next, the first author, with extensive experience in English writing and Knowledge Building, examined the pairs and originating sentence units to identify students' figurative language use and open code contributions for use of figurative devices, as shown in Table 1. In this study, creative ideas were defined as "unique contributions that create epistemic value to the community beyond scientific content."

Results

Table 1 shows five figurative expressions (simile, personification) identified in the discourse, which distributed across the semantic distance scores ranging from 0.22 to 0.61. There were seven additional instances of figurative language that were eliminated due to our effort to exclude repetition of figurative language used in source material. All instances suggest promising input to community discussion. Similes involved *conceptual reframing*—mapping science concepts onto the everyday life domain. Personification reflected students' *perspective-taking through embodiment.*

Table 1

A Summary of Noun Pairs, Semantic Distance, Figurative Devices, and Originating Texts

Noun Pairs	Semantic Distance	Figurative Devices	Originating Texts
light-water	0.6109	Simile	***light*** travels in a ripple waves like when you hit the ***water*** with your hand
light-mirror	0.5465	Personification	***light*** doesn't like the ***mirror*** so it bounces off and makes light travel
light-ball	0.5163	Simile	***light*** acts like a tennis ***ball*** hitting a flat wall
prism-sunlight	0.2326	Personification	when the ***prism*** senses the ***sunlight*** shining on it, the prism cannot stand it anymore
color-paint	0.2217	Simile	even though it wasn't paint, the ***colors*** still mixed liked ***paint*** so it made orange

The rest of the units showed literal use of language and scientific discourse. Noun pair combinations in this group reflected conceptual associations between general words (e.g., sun), familiar light domain words (e.g., shadow), and technical domain terms (e.g., umbra) within the subject domain.

Discussion and next steps

This pilot analysis explored grade 4 students' novel figurative language use to identify promising ideas in knowledge building community discourse. In the current dataset, novel figurative language use, particularly metaphor production, was rare in written scientific discourse of this age group. Some instances of figurative language were excluded from the current analysis because they may have been derived from source material. However, the fact that students reused figurative expressions from source material suggests that the borrowed figurative framing supported their discourse.

Ideas co-occurring with figurative language were distinct from questions and explanations seen in standard discourse with literal language use. The results suggest that students' figurative language use may support them to create a conceptual link across different domains. Similes were used to create authentic, real-life connections across science and everyday concepts, diversifying students' ideas. Personifications may have enabled the students to take an anthropomorphic approach to objectifying, simulating, and discussing complex, abstract concepts.

The methodological implication is that semantic distance between noun pairs may serve as an indicator of conceptual distance students map through similes and metaphors. However, the limitation of Word2Vec in handling polysemous words (e.g., light) will need to be addressed in future research.

The current findings provide only preliminary results of a pilot investigation. As our next step, we plan to expand our analysis to other community datasets, incorporate similarity to resource material, and investigate how ideas that co-occur with figurative language support community discourse.

References

Amabile, T. M. (1988). A model of creativity and innovation in organizations. In B. M. Staw, & L. L. Cummings (Eds.), *Research in organizational behavior* (Vol. 10, pp. 123-167). JAI Press.

Beaty, R., & Johnson, D. (2021). Automating creativity assessment with SemDis: An open platform for computing semantic distance. *Behavior Research Methods, 53*, 757-780. https://doi.org/10.3758/s13428-020-01453-w.

Cameron, L. (2002). Metaphors in the learning of science: A discourse focus. *British Educational Research Journal, 28*, 673-688. https://doi.org/10.1080/0141192022000015534.

Fauconnier, G., & Turner, M. (2003). Conceptual blending, form and meaning. *Recherches En Communication, 19*. https://doi.org/10.14428/rec.v19i19.48413

Gardner, H., Kircher, M., Winner, E., & Perkins, D. (1975). Children's metaphoric productions and preferences. *Journal of Child Language, 2*(1), 125–141. https://doi.org/10.1017/S0305000900000921

Glăveanu, V. P. (2012). What can be done with an egg? Creativity, material objects, and the theory of affordances. *The Journal of Creative Behavior, 46*(3), 192–208. https://doi.org/10.1002/jocb.13

Khanlari, A., Zhu, G., & Scardamalia, M. (2019). Knowledge building analytics to explore crossing disciplinary and grade-level boundaries. *Journal of Learning Analytics, 6*(3), 60–75. https://doi.org/10.18608/jla.2019.63.9

Lakoff, G., & Johnson, M. (1980). The metaphorical structure of the human conceptual system. *Cognitive Science, 4*(2), 195–208. https://doi.org/10.1207/s15516709cog0402_4

Liu, C., Ren, Z., Zhuang, K., He, L., Yan, T., Zeng, R., & Qiu, J. (2020). Semantic association ability mediates the relationship between brain structure and human creativity. *Neuropsychologia, 151*. https://doi.org/10.1016/j.neuropsychologia.2020.107722.

Mednick, S. A. (1962). The associative basis of the creative process. *Psychological Review, 69*(3), 220–232. https://doi.org/10.1037/h0048850

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv:1301.3781v3 [cs.CL].

Piccardo, E. (2017). Plurilingualism as a catalyst for creativity in superdiverse societies: A systemic analysis. Frontiers in Psychology, 8, Article 2169. https://doi.org/10.3389/fpsyg.2017.02169

Pollio, M. R., & Pollio, H. R. (1974). The development of figurative language in children. Journal of *Psycholinguistic Research* 3, 185–201. https://doi.org/10.1007/BF01069237

Resendes, M., Scardamalia, M., Bereiter, C. *et al.* Group-level formative feedback and metadiscourse. *Intern. J. Comput.-Support. Collab. Learn* 10, 309–336 (2015). https://doi.org/10.1007/s11412-015-9219-x

Runco, M. A., & Jaeger, G. J. (2012). The standard definition of creativity. Creativity Research Journal, 24, 92–96. https://doi.org/10.1080/10400419.2012.650092

Thomas, D.P., Walton, J., Tann, K., Shoecraft, K., Lowien, N., & Scott, A. (2025). Figurative language and narrative writing: Insights from high-achieving primary school students. *The Australian Educational Researcher*. https://doi.org/10.1007/s13384-025-00890-w

Yang, T., Zheng, X., Dong, Z., Zheng, Y., & Hou, Y. (2025). Bridging the gap through common features: The semantic structure of creative metaphors. *Psychology of Aesthetics, Creativity, and the Arts.* Advance online publication. https://doi.org/10.1037/aca0000737

International Society of the Learning Sciences

Augmentation, Enhancement, or Replacement? Classifying Generative AI Use in Workforce Development Practice

Jacob Hale, Tessa Forshaw, Mari Longmire
jhale@hks.harvard.edu, tforshaw@fas.harvard.edu, mlongmire@gse.harvard.edu
Graduate School of Education, Harvard University

Abstract: This study examines how frontline workforce development practitioners use GenAI. A survey of 181 practitioners from seven organizations revealed use cases concentrated in learning and instruction (40%), communication and design (39%), and research and analysis (36%). A typology classifying use cases as augmentation (32%), enhancement (26%), replacement (23%), or unspecified (18%) shows that a majority reflect augmentation or enhancement. These results can inform practical guidelines that optimize human-AI collaboration while minimizing replacement of critical thinking.

Introduction and literature review

OpenAI's ChatGPT ushered in a new era of technological advancement when it was introduced in November 2022. Generative artificial intelligence (GenAI) and its expanding ecosystem of tools have begun reshaping how people work, learn, and live, offering benefits such as increased productivity and access to information (Noy & Zhang, 2023). At the same time, concerns have emerged around cognitive offloading, dependency, and harm to learning (Bastani et al., 2024; Cheung, 2025; Kosmyna et al., 2025). While a growing body of research examines GenAI in formal K–12 and higher education settings (Cao & Dede, 2023; Yan et al., 2023), far less attention has been paid to frontline workforce development practitioners—those in community colleges, public workforce agencies, apprenticeship programs, and skills bootcamps—despite their central role in preparing individuals for a changing labor market (Ogunleye et al., 2024).

Understanding practitioner use of GenAI in these contexts is critical because workforce development serves diverse populations—including career changers, jobseekers, and individuals from underrepresented communities—whose instructors and coaches often implement AI-enabled practice with minimal guidance. This paper advances a practice-grounded account of GenAI use in workforce development by asking: (1) how frontline practitioners currently use GenAI across tasks; (2) to what extent those use cases represent augmentation, enhancement, or replacement; and (3) what implications these patterns have for GenAI integration. Drawing on a mixed-methods survey of 181 practitioners across seven organizations, we develop a functional typology and discuss implications for learning design, equity, and policy.

Methodology

We developed an anonymous online survey administered via Qualtrics consisting of ten items: two screening questions, one multiple-choice item on use frequency, and seven open-ended questions informed by existing literature on GenAI adoption (Yan et al., 2023; Ogunleye et al., 2024). The full instrument is available upon request.

Through the researchers' professional network, seven workforce development organizations were invited to distribute the survey to frontline practitioners. These organizations represented a range of contexts: five nonprofits spanning professional certification, career services, clean technology training, labor union apprenticeships, and IT skills development; one regional workforce agency; and one metropolitan workforce development board. They were geographically distributed across the United States and served diverse populations including jobseekers, career changers, apprentices, young adults, and military-connected individuals. Participation was voluntary, and the study received IRB approval prior to data collection. The survey yielded 181 responses, of which 44 were excluded for ineligibility, leaving 137 qualified respondents.

A thematic analysis was conducted following Naeem et al. (2023): two coder-researchers developed a codebook inductively after reviewing a sample of responses, iterating until reliability was validated. We then extended the analysis abductively (Tavory & Timmermans, 2014), posing a theory-driven question: What kind of work is GenAI doing in these contexts—augmenting, enhancing, or replacing practitioner tasks? This cycle yielded the typology in Table 1. The key distinction between augmentation and enhancement is whether the practitioner could complete the task independently: augmentation involves AI supporting tasks already within the practitioner's capability (e.g., drafting a document faster), whereas enhancement involves AI enabling tasks that exceed the practitioner's existing skill set (e.g., correcting grammar for a non-native English speaker). Frequencies

 © ISLS

were analyzed at the response level so that multiple mentions of a single code by one respondent counted as a single instance.

Table 1
Typological Framework for Generative AI Use Cases

Typology	Generative AI's Role	Quotes
Augmentation	To support or accelerate tasks the user can already perform independently	It saves time for workforce professionals by helping draft documents, training materials, or client plans faster.
Enhancement	To enable the user to complete tasks beyond their existing skill level	Individuals with less background in writing and editing benefit from grammatical errors and sentence structure.
Replacement	To replace the user entirely in the completion of the task	AI can respond to any series of questions or answer questions without having to go to a live person

Results

Analysis of 137 qualified responses revealed three primary use case categories: Learning and Instruction (40% of respondents), Communication and Design (39%), and Research and Analysis (36%); percentages exceed 100% because some respondents reported multiple uses. Learning and Instruction encompassed lesson planning, personalized materials, and assessment design. Communication and Design included refining written communications and creating presentations, website content, and images. Research and Analysis involved information synthesis, simplification of complex topics, and data summaries.

Applying the typology, most use cases reflected augmentation (32%), followed by enhancement (26%) and replacement (23%); 18% were unclassified. Overall, 58% of uses fell into augmentation or enhancement. When examined by category, Communication and Design showed the highest rate of replacement (24.5%), while Learning and Instruction leaned toward enhancement (53.5%). At the respondent level, 61.3% reported at least one augmentative use, 45.3% at least one enhancement, and 41.6% at least one replacement—suggesting that most practitioners engage GenAI fluidly across modes rather than being locked into a single pattern.

Discussion and conclusion

Across use cases, most GenAI activity centers on extending or refining existing tasks rather than replacing them, with 58% of uses falling into augmentation or enhancement. Replacement appears mainly in lower-stakes communication and design work—formatting slides, editing text, generating visuals—while the deeper cognitive work of instructional design, assessment, and personalization remains with practitioners. One reason this pattern may emerge is that workforce development tasks vary in the degree of professional judgment they require. Formatting and polishing are procedural and well-suited to delegation, whereas designing a lesson for a specific learner population or evaluating assessment validity demands contextual knowledge that practitioners are reluctant—or unable—to hand off. This suggests that the distribution is not arbitrary but reflects an implicit triage: practitioners delegate where the cost of error is low and retain control where pedagogical or relational stakes are high.

The findings rely on a convenience sample and self-reported data, limiting generalizability, and the study does not assess downstream learning or workforce outcomes. Future research should use in-situ observations and outcome measures to determine whether and how GenAI use translates into measurable learning and workforce gains.

References

Al-Kfairy, M., et al. (2024). Ethical Challenges and Solutions of Generative AI. Informatics, 11, 58.
Bastani, H., et al. (2024). Generative AI can harm learning. SSRN.
Cao, L., & Dede, C. (2023). Navigating a world of generative AI. Harvard Graduate School of Education.
Cheung, M. (2025). From answer machines to ignorant co-learners. CHI '25 Workshop. ACM.
Kosmyna, N., et al. (2025). Your brain on ChatGPT. arXiv.
Naeem, M., et al. (2023). A step-by-step process of thematic analysis. Int. J. Qualitative Methods, 22.

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative AI. Science, 381, 187–192.
Ogunleye, B., et al. (2024). A systematic review of generative AI for teaching and learning. Educ. Sci., 14(6), 636.
Tavory, I., & Timmermans, S. (2014). Abductive analysis. University of Chicago Press.
Yan, L., et al. (2023). Practical and ethical challenges of large language models in education. Br. J. Educ. Technol., 55, 90–112.

 © ISLS

International Society of the Learning Sciences

Investigating Collaborative Weaving Design as a Scaffold for AI Literacy

Carolyn Rosé, Carnegie Mellon University, cp3a@andrew.cmu.edu
Olivia Robinson, Carnegie Mellon University, orobinso@andrew.cmu.edu
Melisa Orta-Martinez, Carnegie Mellon University, mortamar@andrew.cmu.edu
Joey Huang, North Carolina State University, joeyhuang@ncsu.edu
Kylie Peppler, University of California at Irvine, kpeppler@uci.edu

Abstract: This poster presents version 1.0 of a curriculum unit designed to expose students to machine learning concepts related to representation and reasoning, an understudied area in AI literacy. Embedded within a textile art course, this three-week pilot explores whether hands-on weaving activities scaffold or distract from AI concept acquisition. Students design weavable potholders while engaging with an AI agent that demonstrates both capabilities and limitations of large language models.

Introduction

As generative AI technologies become increasingly embedded in daily life—from chatbots to recommender systems—students must develop the ability to critically evaluate these systems, understand their representational logic, and assess their societal impacts (Willige, 2024; Kafai et al., 2024; Kim et al., 2024). Despite increasing exposure to AI tools, even most undergraduates—particularly in non-STEM disciplines—lack this foundational understanding of how AI models represent and manipulate information, what their limitations are, and when human reasoning must take precedence (Veldhuis et al., 2025). Aligned with US AI4K12 (Touretzky et al., 2019) and other similar international efforts (UNESCO, 2022), curricular development and research in AI literacy has aimed to begin infusion of understanding and insight related to AI throughout K-16 education (Veldhuis et al., 2025). Many current AI literacy efforts focus on teaching how to use, configure, or even build AI-enabled services, tools, appliances, and toys (Kafai et al., 2024; Kim et al., 2024). However, another critical need is insight into what is reasonable to expect from current and future AI capabilities, and relatedly, what the continuing critical role of uniquely human capabilities will be (Kaspersen et al., 2021). This level of insight requires some understanding of what the reasons for the enduring limitations are and why they may not be obvious upon limited experience, or even the everyday experience with AI technologies that is now commonplace.

Curriculum materials

The key set of AI insights that this curriculum research effort focuses on is related to human-AI complementarity: the idea that humans and AI systems rely on fundamentally different strategies for problem solving. In particular, whereas humans excel at reasoning about abstractions and decomposing complex problems into smaller, generalizable steps, AI problem solvers struggle to form and use abstractions. In contrast to their human counterparts, AI systems manage to solve problems that require decomposition by retrieving similar cases from a potentially massive repository of past cases, which they may tailor in a limited way. Relatedly, when facing challenges during problem solving, humans are able to gradually adapt and improve a multi-step plan for a solution, while AI problem solvers are more likely to regenerate a solution from scratch, again by searching for a different prior case. At the surface, both strategies may frequently appear similar, and thus AI problem solvers are observed to provide impressive looking responses to a wide-range of queries. However, as researchers probe more deeply into the inner workings of AI problem solvers, it becomes clear that the apparent increase in capabilities of current AI models is mainly due to the massive increase in the size of both the models and the data repositories used to construct them.

To make these differences concrete, we use weaving on a shaft loom as a physical and conceptual analogy for machine learning representations. Weaving on a shaft loom provides a powerful physical embodiment of machine learning concepts, with the loom physically representing the multiplication of three matrices (threading, tie-up, and treadling). Students perform this multiplication by following weaving drafts, connecting embodied experience to abstract computation (i.e., matrix factorization). Designing weaving patterns further mirrors how machine learning breaks down complex patterns into underlying representations. Students must find the right level of abstraction (i.e., detailed enough to preserve desired structure, but general enough to be woven), mirroring the machine learning tension between overfitting and underfitting, To make designs weavable, students must simplify them to emphasize regular, hierarchical structures in order to use the computational power of the shaft structure

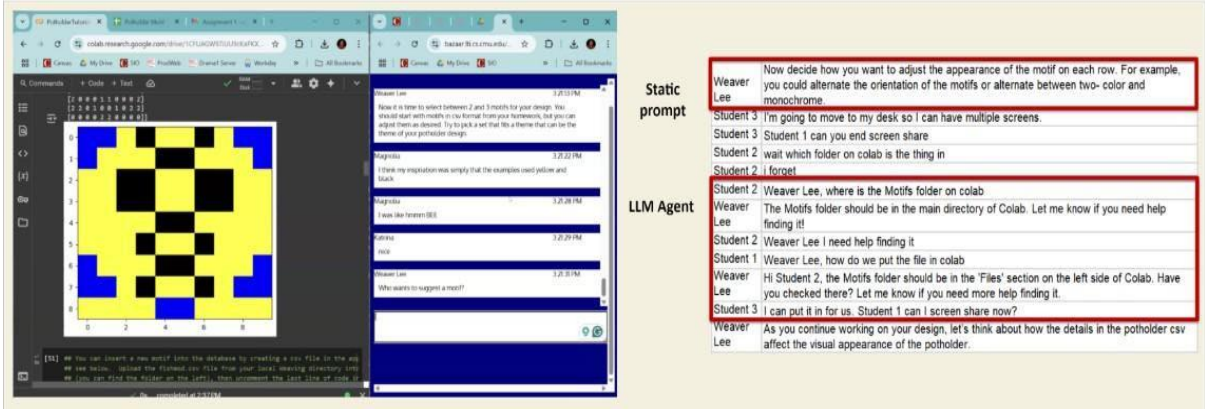
 © ISLS

judiciously. This parallels how ML models compress complex data into lower-dimensional embeddings, often losing information in the process (i.e., compression and approximation). Finally, just as looms place physical limitations on what patterns they can produce, neural networks have representational constraints that limit their capabilities (representational constraints).When designing, students work in a collaborative digital environment featuring a shared Jupyter notebook for interactive programming, integrated chat tool for team communication, and an AI facilitation agent (named "Weaver Lee") that provides both structured guidance and demonstrates LLM limitations in real-time (Figure 1). Weaver Lee functions as both a collaborative design assistant and a critical reflection scaffold. It prompts students to iterate on designs, surfaces contradictions in abstraction, and occasionally fails in subtle ways, prompting metacognitive discussions about the limits of LLMs. It is designed around principles of productive failure and epistemic agency, helping students develop a calibrated understanding of what AI can and cannot do. Just as introduced earlier, the shaft loom's representational limits provide a tangible counterpart to the compression and compositional challenges inherent in neural network representations .

Despite their limitations, LLM agents are able to serve as facilitators of collaborative learning sessions. By focusing on tasks well-matched to LLM capabilities, such as using reflection prompts to support learning in collaborative software development (Naik et al., 2024, 2025), we can more effectively leverage GenAI in practice. We take a similar approach to that taken in Naik's (2024; 2025) past work with AI agent support for collaborative engagement with the activity. The current pilot did not separately analyze the specific impact of the AI agent on learning outcomes.

Figure 1

Collaborative Design Environment (on left) and Example Interaction (on right).



Pilot

A small pilot of the three-week unit was conducted in an undergraduate interdisciplinary textile art course at a mid-Atlantic university. Four teams of students (N= 12) engaged in a potholder design task that required them to work across three representational levels: raw matrix data, loom mechanics, and high-level pattern logic. Students completed a pre-assessment after engaging with the assigned reading and interactive lecture introducing the core concepts. Because the pre-assessment followed this initial instruction, it does not serve as a baseline measure of prior knowledge, which was intentional because the goal was to measure learning specifically during the collaborative design portion of the unit. This, the pre–post comparison reflects any shifts in understanding that occurred during the design and reflection phases. The three-week period of time during which they engaged in the design activity began with individual design work, with a single collaborative session in the middle, and final individual finalization of the design. Overall, the results from the pilot study suggest that during the design and reflection, students moved away from a task focus and towards engagement with AI concepts. However, their written answers demonstrated that they still did not reach mastery. While the quantitative comparison between pretest and posttest focusing on multiple-choice questions demonstrated substantial learning gains, the fact that students continued to struggle with abstract symbolic constraints suggests a need for more structured support in transitioning from concrete activities to analogous abstract reasoning, which will be addressed in subsequent work.

 © ISLS

References

Kim, D. Y. J., Ravi, P., Williams, R., & Yoo, D. (2024). App Planner: Utilizing generative AI in K–12 mobile app development education. In *Proceedings of the 23rd Annual ACM Interaction Design and Children Conference (IDC '24)* (pp. 770–775). https://doi.org/10.1145/3628516.365939

Kaspersen, M. H., Graungaard, D., Bouvin, N. O., Petersen, M. G., & Eriksson, E. (2021). Towards a model of progression in computational empowerment in education. *International Journal of Child-Computer Interaction, 29,* Article 100302. https://doi.org/10.1016/j.ijcci.2021.100302

Naik, A., Yin, J. R., Kamath, A., Ma, Q., Wu, S. T., Murray, R. C., Bogart, C., Sakr, M., & Rosé, C. P. (2024). Generating situated reflection triggers about alternative solution paths: A case study of generative AI for computer-supported collaborative learning. In A. M. Olney, I.-A. Chounta, Z. Liu, O. C. Santos, & I. I. Bittencourt (Eds.), *Artificial Intelligence in Education: 25th International Conference, AIED 2024, Recife, Brazil, July 8–12, 2024, Proceedings, Part I* (Lecture Notes in Computer Science, Vol. 14829, pp. 46–59). Springer.

Naik, A., Yin, J. R., Kamath, A., Ma, Q., Wu, S. T., Murray, R. C., Bogart, C., Sakr, M., & Rosé, C. P. (2025). Providing tailored reflection instructions in collaborative learning using large language models. *British Journal of Educational Technology*, *56*(2), 531–550. https://doi.org/10.1111/bjet.13548

Touretzky, D. S., Gardner-McCune, C., Martin, F., & Seehorn, D. (2019). Envisioning AI for K–12: What should every child know about AI? In *Proceedings of the AAAI Conference on Artificial Intelligence, 33*(1), Article 9795. https://doi.org/10.1609/aaai.v33i01.33019795

UNESCO. (2022). *K–12 AI curricula: A mapping of government-endorsed AI curricula*. https://unesdoc.unesco.org/ark:/48223/pf0000380602

Veldhuis, A., Lo, P. Y., Kenny, S., & Antle, A. N. (2025). Critical artificial intelligence literacy: A scoping review and framework synthesis. *International Journal of Child-Computer Interaction, 43,* Article 10070. https://doi.org/10.1016/j.ijcci.2024.10070

Willige, A. (2024). AI skills are becoming more important than job experience: Here's how job seekers can adapt. *World Economic Forum*. https://www.weforum.org/stories/2024/06/ai-skills-rival-job-experience-in-recruitment/

Acknowledgments

This research was funded in part by NSF grant DUE 2100401.

International Society of the Learning Sciences

Reflective Assessment for Authentic Research among High School Students: Enhancing Epistemic Agency for Knowledge Building

Lei Ding, Jianwei Zhang, Thomas Underwood, Weiting Qi, Dongni Guo, Alae Sherafat
dding3@albany.edu, jzhang1@albany.edu, tunderwood@albany.edu, wqi3@albany.edu, dguo@albany.edu, asherafat@albany.edu
University at Albany, State University of New York, USA

Lana Hower, Jennifer Muirhead, Maryann Dunn
lhower@techvalleyhigh.org, jmuirhead@techvalleyhigh.org, maryann.dunn@techvalleyhigh.org
Tech Valley High School, Albany, NY, USA

Abstract: This study investigates how high school students engage with reflective assessment during their authentic research experience. An individual interview was conducted to explore students' reflections on their use of a reflective assessment document. Grounded theory analysis of the interview data and their reflective assessment document revealed key engagement. As a compelling theme, this document guides students to interpret research evidence and identify progressive questions. The findings inform educational designs to cultivate epistemic agency in a more student-centered, open-ended, collaborative inquiry context.

Introduction

To prepare students for college, careers, and citizenship in the rapidly changing world, education needs to cultivate students' abilities to solve complex problems. In this context, recent educational reform highlights the need to engage students in authentic knowledge creation practices, similar to those used by real-world researchers to explore and address real-life problems (OECD, 2018). The work reported herein explores how such authentic research experiences can be assessed in a way that engages students' agency for collaborative knowledge building.

Our design of authentic research for high school students draws upon frameworks of community-based learning (Rogoff, 2014) and Knowledge Building (Scardamalia, 2002). Students enact epistemic agency for charting their research process in collaboration with their teachers and community partners. Epistemic agency refers to "a capacity that enables groups to deliberately carry out collaborative, knowledge-driven activities to create shared knowledge objects" (Damșa et al., 2010). Thus, students intentionally take epistemic agency and collective responsibility for collective idea improvement while navigating the process of collective inquiry to work towards a shared epistemic artifact.

Implementing such authentic research experience among high school students requires new ways to assess student engagement and progress. Practices of authentic research experience in K-12 engage in predefined courses and tasks (Aydeniz et al., 2011; Melton et al., 2022), limiting student epistemic agency to engage in open-ended inquiry and do the knowledge building. Assessment in these practices usually emphasizes the acquisition of knowledge and skills in individual students, blocking students from working with their peers to co-construct new knowledge and develop competencies. This paper addresses this gap by designing reflective assessments and investigating student engagement with one document of reflective assessment. Reflective assessment is a type of formative assessment that guides students to enact metacognitive processes for charting and reflecting on collaborative inquiry. More research should be done to investigate favorable circumstances in which reflective assessment could enhance students' high-level epistemic engagement in real-world knowledge building work (Scardamalia, 2002; van Aalst & Chan, 2007; Dan & Zhang, 2018). Based on this, our research question investigates how students engage in reflective assessment to address authentic research problems.

Method

Classroom contexts and reflective assessment design

This study is part of an ongoing design-based research project. For the past three years, our research team has been collaborating with a team of teachers and community partners (CPs) to develop a collaborative environment for high school students to engage in authentic research experiences (Zhang et al., 2024, 2025; Ding et al., 2025). At the beginning of the project, our research team shared a few knowledge building principles with students and teachers, including continual idea improvement, collaborative build-on and rise-above reflection (Scardamalia, 2002). Thus, students get an understanding of research. Research is an ever-deepening process, they can build on their peers' contributions to advance the community's knowledge, and they need to chart, reflect on the paths of inquiry together, and search for new goals during this process. As a part of building this collaborative environment, our research team worked with three teachers to develop the first iteration of reflective assessment design to

 © ISLS

enhance students' epistemic agency. To support students' idea improvement, metacognitive engagement, and ongoing reflection, the design mainly includes a reflective assessment document for guiding students over the cycle of setting the shared goal for their research, documenting the individual research activities for solving their research question, team discussion for monitoring and reflecting on their research findings, and the team's next steps. What is more, three teachers were present in the classroom to give instructions about the use of this document. They also monitored the progress of students' teams through the document and provided guidance to address students' needs. Students received feedback from their peers, teachers, and partners during the project.

Data sources and analysis

The data sources are mainly each team's reflective assessment document, known as research tracking document (RTD) by students, and student interviews. A total of 17 students volunteered to participate in an interview about their research experience. The current study zoomed in one interview question focused on students' reflection on RTD: I remember your teachers introduced the Research Tracking Document. Here is your group's doc, you could take a look at it. Did you find it useful for your research? If yes, in what ways? In total, 6 out of 17 students' interviews were excluded because they did not provide sufficient clarity or details to support meaningful sense-making. The interviews were transcribed and analyzed using grounded theory approach (Strauss & Corbin, 1998). Before coding, the first author read and re-read transcriptions of interviews and team's RTD, creating raw codes to capture students' engagement with reflective assessment. Then, for axial coding, the first author reviewed the raw codes and examples to identify thematic clusters. The second coder selected 20% of students' interviews and repeated the process of grounded theory coding. The interrater reliability between two researchers was 95%.

Findings

Theme 1: Eliciting research evidence in an organized way

This theme focused on processing information using RTD. Students emphasized that they use RTD to organize components of research, such as research topics, questions, and sources, more clearly and systematically.

Theme 2: Guiding students to interpret research evidence and identify progressive questions

The second theme focused on the epistemic function in the process of research. Students highlighted that reflective assessment helped them answer the existing research question and uncover emerging questions in their research progress. David (pseudonym) told us: "You do it in a linear thinking process from one to another. So our first thing was like 'W' Community Board. We found that, and then we summarized it, and then what are our next steps, and what did we take away from this? And we just kept doing that, did that like 10 or 15 times."

Theme 3: Developing a coherent understanding of research thinking

The third theme conveyed the epistemic results of using RTD, students developed a coherent understanding of research thinking by using a linear thinking process manifested in RTD. Kelly mentioned that "The document [Research Tracking Document] really helped organize my thoughts, even, like some of it wasn't just research. And I still put my main points of our meetings, like meeting with our CP and other stuff."

Theme 4: Finding the ways to solve their challenges

The fourth theme focused on epistemic results of RTD. In the project, students needed to solve the challenges raised by the CPs. Students noticed that using RTD to do research would help them find their way to solve their challenges. As we heard from Leo (pseudonym), "We go down the line of questions until we are good on that, keep going until you find the answer you need. All of the answers work together [to solve our challenge]."

Theme 5: Monitoring and reflecting on research to have a clear sense of progress

The fifth theme focused on the regulative function of RTD. Students highlighted the function of monitoring, reflecting their research activities, thus having a clear sense of research progress. As Max (pseudonym) told us, "This [RTD] helped pace our research. This helped us know where we were at and when we were done."

Discussion

Essential to students' epistemic agency, the findings shed light on students' engagement with reflective assessment. The study is among the first to design and implement the reflective assessment for high school students in authentic research experience, integral to open-ended collaborative inquiry. The findings indicate that reflective assessment is conducive to the development of epistemic and regulative dimensions of epistemic agency.

References

Aydeniz, M., Baksa, K., & Skinner, J. (2011). Understanding the impact of an apprenticeship-based scientific research program on high school students' understanding of scientific inquiry. Journal of Science Education and Technology, 20(4), 403-421.

Damşa, C. I. et al. (2010). Shared epistemic agency: An empirical study of an emergent construct. *Journal of the Learning Sciences, 19*, 143–186.

Ding, L., Sherafat, A., Zhang, J., Underwood, T., Guo, D., Muirhead, J., Hower, L., & Kerr, J. (2025). Cultivate a research mindset: How do high school students understand and approach authentic research? In J. Oshima, B. Chen, F. Vogel, & J. Järvelä (Eds.), *Proceedings of the 18th International Conference on Computer-Supported Collaborative Learning* (CSCL 2025) (pp. 178–186). International Society of the Learning Sciences. https://doi.org/10.22318/cscl2025.829130

Jan Van Aalst, J., & Chan, C.K. (2007). Student-Directed Assessment of Knowledge Building Using Electronic Portfolios. Journal of the Learning Sciences, 16, 175-220.

Melton, J.W., Saiful, J.A., & Shein, P.P. (2022). Interdisciplinary STEM program on authentic aerosol science research and students' systems thinking approach in problem-solving. *International Journal of Science Education, 44*, 1419-1439.

Rogoff, B. (2014). Learning by Observing and Pitching In to Family and Community Endeavors: An Orientation. Human Development, 57(2–3), 69–81.

Organization for Economic Cooperation and Development (OECD). (2018). *The future of education and skills: Education 2030*. OECD Education Working Papers.

Strauss, A. L., & Corbin, J. M. (Eds.). (1997). *Grounded theory in practice*. Sage Publications, Inc.

Schwartz, D. L., Lindgren, R., & Lewis, S. (2009). Constructivism in an age of non-constructivist assessments. In S. Tobias & T. M. Duffy (Eds.), *Constructivist instruction: Success or failure?* (pp. 34–61). Routledge/Taylor & Francis Group.

Scardamalia, M. (2002). Collective cognitive responsibility for the advancement of knowledge. In B. Smith (Ed.), *Liberal education in a knowledge society* (pp. 76–98). Open Court.

Tao, D., & Zhang, J. (2018). Forming shared inquiry structures to support knowledge building in a grade 5 community. *Instructional Science, 46*, 563-592.

Zhang, J., Song, G., Underwood, T., Guo, D., Espinosa, K., Hower, L., Muirhead, J., & Kerr, J. (2024). Cultivate research literacy: High school students work with community partners to investigate real-world challenges. In R. Lindgren, T. I. Asino, E. A. Kyza, C. K. Looi, D. T. Keifert, & E. Suárez (Eds.), *Proceedings of the 18th International Conference of the Learning Sciences* (ICLS 2024) (pp. 690–697). International Society of the Learning Sciences. https://doi.org/10.22318/icls2024.111297

Zhang, J., Sun, Y., & Underwood, T. (2025). "Research is like a wild animal": What do student-generated metaphors reveal about their understanding of research? In A. Rajala, A. Cortez, H. Hofmann, A. Jornet, H. Lotz-Sisitka, & L. Markauskaite (Eds.), *Proceedings of the 19th International Conference of the Learning Sciences* (ICLS 2025) (pp. 1953–1957). International Society of the Learning Sciences. https://doi.org/10.22318/icls2025.403402

Acknowledgments

We would like to thank the students, community partners, and teachers for their creative work and support, enabling this research.

International Society of the Learning Sciences

Using Generative AI as an Epistemic Catalyst for Collaborative Lesson Planning

Yuxin Chen, Drew Gossen
yuxinchen@southalabama.edu, dgossen@southalabama.edu
University of South Alabama

Abstract: This study examines the integration of generative AI to cultivate critical AI literacy and epistemic understanding in pre-service teachers. The AI-generated lesson plan is conceptualized as a boundary object, compelling collaborative groups to negotiate and externalize their disciplinary knowledge and reasoning, considering the 5E inquiry model and discourse structure. A multi-modal analysis is utilized to quantify and qualify shifts in epistemic engagement, serving as a robust testbed for necessary design principles in critical collaboration with intelligent systems.

Introduction

The rapid and ubiquitous integration of generative artificial intelligence (e.g., ChatGPT) fundamentally alters prevailing educational praxis, thereby necessitating investigative research that moves beyond the mere adoption of instrumental tools (Holmes & Tuomi, 2022; Cress & Kimmerle, 2023). This transformation holds particular salience within teacher education, as the preparation of pre-service teachers (PSTs) must necessitate the critical evaluation and integration of AI outputs in a manner deemed pedagogically sound (Long & Magerko, 2020; Guan et al., 2025). Current scholarly discourse frequently positions artificial intelligence as a utility for the attainment of automation or efficiency, often neglecting its capacity to reshape the dynamics of collaborative learning and the evolution of professional epistemic understanding (Chen 2025; Wu et al., 2025). The growing influence of generative AI reshapes the locus of epistemic authority in education, challenging traditional grounds for knowledge and underscoring the need to reaffirm human-centered agency (Peters & Dobber, 2025).

Despite the substantial promise of artificial intelligence, its theoretical function in the domain of collaborative learning remains nascent. Much existing research positions AI as a tool for optimizing learning efficiency and performance, rather than as an active participant in the social and epistemic processes of collaboration. However, it remains unclear how AI not only simply support existing collaborative practices but also actively reshape the social and cognitive dimensions of group work (Järvelä et al., 2023). Two gaps remain in the literature on AI in collaborative learning. First, although AI literacy encompasses both technical skills and the critical capacities to evaluate and responsibly use AI outputs (Long & Magerko, 2020), there is limited empirical research on how AI is mediating learners' epistemic understanding in CSCL contexts (Cress & Kimmerle, 2023; Cheung et al., 2025). Second, more empirical research is required to inform how to leverage AI for future educators by designing and examining the collaborative interaction in teacher-AI collaboration (Holstein et al., 2019; Kim, 2024).

This study presents the design of an intervention aimed at fostering critical AI literacy and deeper epistemic understanding among PSTs in a collaborative lesson planning context. The intervention is grounded in constructivist theory, which posits that knowledge is actively co-constructed through structured interaction. The pedagogical core employs the 5E instructional model (Duran & Duran, 2004), an inquiry-based framework central to science education, integrated with Accountable Talk strategies (Michaels et al., 2008), which promote scientific discourse among teachers. This integrated pedagogical framework structures collaborative lesson planning to promote the co-construction of knowledge, generate accountable talk discourse moves, and advance group regulatory skills. This design reconceptualizes the role of artificial intelligence in collaborative learning as a socio-epistemic agent whose outputs are to be examined, debated, and reworked through group dialogue. By inviting participants to negotiate rather than accept AI-generated suggestions, the study foregrounds the relational and epistemic dimensions of human–AI co-construction, specifically addressing two research questions (RQ):

RQ1. How does AI-assisted collaboration influence PSTs' epistemic understanding of lesson planning

RQ2. How do PSTs interpret and critique AI suggestions during discussions?

Research design

The intervention employs a within-subject, comparative design implemented in a university-level science teacher education course. This structure enables the systematic observation of epistemic change by examining identical groups across two distinct conditions. In Phase 1 (Lesson Plan 1; LP1)**,** groups collaboratively revise lesson plans without the use of generative AI. The focus is solely on applying the 5E instructional model and Accountable

 © ISLS

Talk strategies to establish a pedagogical baseline. This phase provides essential diagnostic insight into participants' existing pedagogical reasoning and discourse patterns. In Phase 2 (Lesson Plan 2; LP2**)**, the same groups are required to engage with generative AI (ChatGPT). Participants use the AI to generate lesson components or full drafts, which are then collectively critiqued, negotiated, and refined in alignment with the 5E and Accountable Talk frameworks. Within this design, AI-generated content in LP2 functions as a boundary object, an artifact situated at the intersection of human pedagogical reasoning and machine-generated content (Star & Griesemer, 1989). Recent research recognizes generative AI as a cognitive collaborator that enhances boundary objects by mediating shared understanding (Naganuma et al., 2024). Specifically, the AI output functions as a Partially Shared Object (PSO) that allows collaborators to coordinate their work while maintaining distinct interpretive perspectives (Akkerman & Bakker, 2011; Naganuma et al., 2024). By encouraging participants to articulate and justify their acceptance, rejection, or transformation of AI contributions, the design externalizes pedagogical knowledge and strengthens PSTs' epistemic commitments.

Methods

To capture the multi-layered nature of AI-enhanced collaboration, this study employs a multi-modal mixed-methods design (see Table 1) design triangulating data from human-ai interaction, group collaboration, and physical artifacts. Two primary instruments were utilized to assess PST's epistemic understanding and socially shared regulation. A pre–post test measured conceptual growth regarding inquiry-based learning, the 5E instructional model, and scientific discourse, combining scenario-based reasoning and concept-mapping tasks to capture how participants articulate, justify, and apply 5E cycles and AT principles (Crawford & Capps, 2018). Complementing this, the Adaptive Instrument for Regulation of Emotions (AIRE) was adapted to the teacher education context to assess socially shared regulation during collaborative lesson planning (Järvenoja et al., 2013). Situated within this context, the survey captures how PSTs monitor, express, and co-regulate affective and motivational states within their groups, providing insight into both individual experiences and group-level regulation. Together, these instruments address RQ1, illuminating how PSTs develop epistemic understanding and socially shared regulation in AI-supported collaboration.

To address RQ2 (Critical Engagement with AI)**,** analysis focuses on the socio-technical dynamics emerging across individual and group levels. Individual ChatGPT trace data are examined through Lag Sequential Analysis (LSA) to model the temporal dependencies within AI-mediated interactions, identifying recurrent behavioral patterns such as prompting, critique, and reformulation (Bakeman & Gottman, 1997). This sequential modeling reveals how PSTs critically engage with and regulate AI input during lesson planning. Complementarily, Interaction Analysis (IA) is applied to the group discussion videos, focusing on how participants collectively evaluate, challenge, or integrate AI-generated content within collaborative dialogue (Jordan & Henderson, 1995). This approach foregrounds the relational processes through which AI suggestions become objects of shared reasoning and negotiation (Strat & Griesemer, 1989).

Table 1

Data Sources and Analysis Methods

	Primary Data Sources	Analysis Methods
RQ1	Pre–Post Epistemic Understanding Test; Adapted AIRE Survey; Final Lesson Plans	Quantitative: Paired *t*-tests (Pre vs. Post; LP1 vs. LP2), Cohen's *d*; correlations. Qualitative: Open-ended responses for epistemic understanding changes.
RQ2	Group Discussion Videos; ChatGPT Trace Data	LSA to model sequences in human-ai interaction; IA to examine nuanced discourse around AI use.

Significance

This AI-enhanced collaborative lesson planning design positions PSTs as active agents in evaluating, integrating, and adapting AI suggestions, fostering ownership, collaborative reasoning, and deeper epistemic understanding. By centering PSTs in the learning process, the intervention promotes reflective and socially shared regulation rather than simple task completion. AI output is treated as a boundary object, mediating interactions and supporting the articulation and refinement of disciplinary knowledge and reasoning. The design also provides a methodological model for examining AI-mediated collaboration, combining Interaction Analysis and Lag Sequential Analysis to capture both individual and group-level engagement. Future research can investigate impacts on lesson plan quality, group discourse, and AI literacy development, as well as explore diverse tools, scaffolds, and collaborative formats to optimize AI integration in teacher education.

 © ISLS

References

Akkerman, S. F., & Bakker, A. (2011). Boundary crossing and boundary objects. *Review of Educational Research, 81*(2), 132–169.

Bakeman, R., & Gottman, J. M. (1997). *Observing interaction: An introduction to sequential analysis*. Cambridge university press.

Cheung, K. K. C., Long, Y., Liu, Q., & Chan, H. Y. (2025). Unpacking epistemic insights of artificial intelligence (AI) in science education: A systematic review. *Science & Education*, *34*(2), 747-777.

Chen, B. (2025). Beyond tools: Generative AI as epistemic infrastructure in education. *arXiv preprint arXiv:2504.06928*.

Crawford, B. A., & Capps, D. K. (2018). *Teaching and learning science as inquiry*. In N. G. Lederman, D. L. Zeidler, & J. S. Lederman (Eds.), *Handbook of research on science education* (Vol. 2). Routledge.

Cress, U., & Kimmerle, J. (2023). Co-constructing knowledge with generative AI tools: Reflections from a CSCL perspective. *International Journal of Computer-Supported Collaborative Learning*, *18*(4), 607-614.

Duran, E., Duran, L., Haney, J., & Scheuermann, A. (2011). A learning cycle for all students. *The Science Teacher*, *78*(3), 56.

Guan, L., Zhang, Y., & Gu, M. M. (2025). Pre-service teachers preparedness for AI-integrated education: An investigation from perceptions, capabilities, and teachers' identity changes. *Computers and Education: Artificial Intelligence*, *8*, 100341.

Holmes, W., & Tuomi, I. (2022). State of the art and practice in AI in education. *European journal of education*, *57*(4), 542-570.

Holstein, K., McLaren, B. M., & Aleven, V. (2019, June). Designing for complementarity: Teacher and student needs for orchestration support in AI-enhanced classrooms. In *International conference on artificial intelligence in education* (pp. 157-171). Cham: Springer International Publishing.

Järvelä, S., Fischer, F., & Suthers, D. (2023). Socio-technical dynamics and challenges in AI-enhanced collaborative learning. *International Journal of Computer-Supported Collaborative Learning, 18*(1), 1–15.

Jordan, B., & Henderson, A. (1995). Interaction analysis: Foundations and practice. *The Journal of the Learning Sciences, 4*(1), 39–103.

Jose, B., Cleetus, A., Joseph, B., Joseph, L., Jose, B., & John, A. K. (2025, August). Epistemic authority and generative AI in learning spaces: rethinking knowledge in the algorithmic age. In *Frontiers in Education* (Vol. 10, p. 1647687). Frontiers.

Kim, J. (2024). Leading teachers' perspective on teacher-AI collaboration in education. *Education and information technologies*, *29*(7), 8693-8724.

Long, D., & Magerko, B. (2020, April). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1-16).

Michaels, S., O'Connor, C., & Resnick, L. B. (2008). *Talk science: A professional development program for the reform of classroom discourse*. National Academy Press.

Naganuma, S., Minematsu, T., Matsueda, K., & Oshima, J. (2024). Use of generative AI for boundary crossing in interdisciplinary collaborative research. In *Proceedings of the 16th International Conference on Computer-Supported Collaborative Learning (CSCL 2024)* (pp. 67–74). International Society of the Learning Sciences (ISLS).

Star, S. L., & Griesemer, J. R. (1989). Institutional ecology, 'translations' and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social Studies of Science, 19*(3), 387–422.

Wu, J. Y., Lee, Y. H., Chai, C. S., & Tsai, C. C. (2025). Strengthening human epistemic agency in the symbiotic learning partnership with generative artificial intelligence. *Educational Researcher, 54*(6), 401–414.

International Society of the Learning Sciences

Assessing Student Collaboration with Multimodal Data and Large Language Models

Shan Zhang, University of Florida, zhangshan@ufl.edu
Hongming Li, University of Florida, hli3@ufl.edu
Seiyon Lee, University of Florida, leeseiyon@ufl.edu
Ji-Eun Lee, Singapore University of Technology and Design, jieun_lee@sutd.edu.sg
Noah L. Schroeder, University of Florida, schroedern@ufl.edu
Anthony F. Botelho, University of Florida, abotelho@coe.ufl.edu

Abstract: We introduce a multimodal evaluation approach that leverages large language models to assess group collaboration quality in computer-supported collaborative learning environments. Using synchronized log data, dialogue transcripts, and human-coded annotations from undergraduate dyads working in the Graspable Math platform, we develop a theoretically grounded rubric capturing cognitive, social, and regulatory dimensions of collaboration, and examine the extent to which LLM-generated evaluations align with expert human judgments.

Introduction

Collaborative learning (CL) supports deeper understanding through shared reasoning and knowledge co-construction (Dillenbourg, 1999). In computer-supported collaborative learning (CSCL), researchers increasingly rely on multimodal data (e.g., logs, dialogue) to capture these complex processes (Sharma & Giannakos, 2020). However, evaluating collaboration quality remains challenging, as rubric-based approaches require extensive human coding and are difficult to scale (Lee et al., 2025). Recent advances in large language models (LLMs) offer a promising alternative for automating the analysis of unstructured educational data, including dialogue and log traces (Liu et al., 2024). While prior work has shown that LLMs can approximate human coding for well-defined constructs, their ability to evaluate group-level collaboration quality using multimodal inputs remains underexplored. To address this gap, this study explores how LLMs can be leveraged to scale the evaluation of pair collaboration based on log data, dialogue data, and human-coded annotations. Specifically: 1) RQ1:What aspects of collaboration quality should be considered when developing a group-level collaboration rubric grounded in the theoretical framework? 2) RQ2: How does the LLM perform in grading group collaboration quality using combined input data (logs, dialogue, and precoded annotations)?

Methods

This study involved 22 undergraduate students (11 dyads) who collaboratively solved algebra problems using the Graspable Math platform (see Figure 1). We recorded both log and dialogue data, providing synchronized multimodal records of student interactions. To address RQ1, we developed a theoretically grounded collaboration rubric based on prior frameworks, Mercer's Exploratory Talk Framework (1995), Weinberger and Fischer's (2006) Argumentative Knowledge Construction model, and Järvelä and Hadwin's (2013) Socially Shared Regulation of Learning perspective, capturing cognitive, social, and regulatory dimensions of collaboration. Dialogue transcripts were manually coded using an established scheme (Zhang et al., 2024) with strong inter-rater reliability ($\kappa > 0.75$), and log data were processed into structured, time-aligned action sequences. Details about dialogue, and log coding schemes are available on the Open Science Framework: https://osf.io/mnxts/overview?view_only=6c0769b0bfcb44998b5a29b6eb184f42. To address RQ2, we constructed multimodal inputs combining logs, dialogue, and human-coded annotations, and evaluated LLM performance (e.g., GPT-5) in grading collaboration quality against expert ratings. Model outputs were assessed using macro F1 and Quadratic Weighted Kappa (QWK).

 © ISLS

Figure 1
(a) GM's Collaborative Whiteboard Interface (a) Algebraic Problem Solving Process(b)



(a) (b)

Results

The rubric in Table 1 comprised six dimensions capturing key aspects of effective collaboration: Participation Balance, Responsiveness & Listening, Idea Elaboration & Co-construction, Questioning & Clarification, Conflict Resolution & Negotiation, and On-task Focus & Shared Goals. These six dimensions were organized into cognitive, social, and regulatory aspects, reflecting the multifaceted nature of collaborative learning. Each dimension was rated on a three-point scale, with an "NA" option when not applicable. Two coders jointly reviewed 20% of transcripts and video recordings to calibrate their understanding of the rubric. Inter-rater reliability was established at 0.84 QWK, after which one coder completed the remaining ratings. After scoring all 11 groups across their attempted problems (n = 49), we computed a total collaboration score per problem by summing ratings across all six dimensions and dividing by the number of applicable dimensions (excluding any marked "NA"). For RQ2, using these expert ratings as ground truth, we evaluated LLM performance under the input data of log, dialogue, and precoded. GPT-5 achieved an F1 score of 0.46 and a QWK of 0.31. This suggests that GPT-5 demonstrates some ability to clarify collaboration quality, but its alignment with expert judgments is still relatively modest. Our findings contribute to the ongoing discourse of using AI for qualitative analyses and provide evidence on the extent to which LLMs can approximate expert evaluations of collaborative processes.

Table 1
Collaboration Quality Rubric

Dimensions	High (3)	Moderate (2)	Low (1)
Participation Balance	Equal contribution; ideas co-developed; turn-taking	Few dominant voices, others still contribute	One member dominates; others passive or silent
Responsiveness & Listening	Active listening; ideas built on and referenced	Some acknowledgment; limited topic building	Ideas ignored; no clear listening; disconnection
Idea Elaboration & Co-construction	Members elaborate, reason, and co-construct meaning	Some elaboration, mostly individual	Simple statements; little or no joint meaning-making
Questioning & Clarification	Frequent open-ended questions; collaborative clarification	Some simple or rhetorical questions; limited follow-up	Rare questioning; minimal clarification
Conflict Resolution & Negotiation	Differences explored respectfully; shared decisions reached	Conflicts noted but not fully resolved	Conflict avoided, suppressed, or escalated
On-task Focus & Shared Goals	Fully on-task; clear shared goals and collaboration	Mostly on-task; some distraction	Off-task; little evidence of shared goals or joint effort

Note. Each dimension includes an "N/A" option. Descriptions are abbreviated here due to space limitations

 © ISLS

References

Dillenbourg, P. (1999). *What do you mean by "collaborative learning"?*

Mercer, N. (1995). *The Guided Construction of Knowledge: Talk Amongst Teachers and Learners*. Multilingual Matters.

Sharma, K., & Giannakos, M. (2020). Multimodal data capabilities for learning: What can multimodal data tell us about learning? *British Journal of Educational Technology*, *51*(5), 1450–1484. https://doi.org/10.1111/bjet.12993

Järvelä, S., Järvenoja, H., Malmberg, J., & Hadwin, A. F. (2013). Exploring Socially Shared Regulation in the Context of Collaboration. *Journal of Cognitive Education and Psychology*, *12*(3), 267–286.

Lee, Y., Kwon, C., Seoh, S., Gweon, G., Stamper, J., & Rosé, C. P. (2025). Capturing Collaborative Competency with GPT-4o and ENA. In *Proceedings of the 18th International Conference on Computer-Supported Collaborative Learning-CSCL 2025, pp. 72-80*. International Society of the Learning Sciences.

Liu, X., Zhang, J., Barany, A., Pankiewicz, M., & Baker, R. S. (2024). Assessing the Potential and Limits of Large Language Models in Qualitative Coding. In Y. J. Kim & Z. Swiecki (Eds.), *Advances in Quantitative Ethnography* (pp. 89–103). Springer Nature Switzerland.

Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & Education*, *46*(1), 71–95.

Zhang, S., Earle-Randell, T. V., Shen, Q., Botelho, A. F., Israel, M., Boyer, K. E., ... & WIEBE, E. (2024, November). Predicting and Analyzing Students' Higher-Order Questions in Collaborative Problem-Solving. In *International Conference on Computers in Education*. APSCE.

Acknowledgements

We would like to thank the National Science Foundation (#2331379), Institute of Education Sciences (#R305B230007), Gates Foundation (#078981), Support from the Learning Engineering Tools Competition, and other anonymous philanthropy.

International Society of the Learning Sciences

A Teacher-in-the-Loop Multi-Agent Framework for Transparent Essay Scoring and Feedback

Ali Keramati, Mark Warschauer
a.kera@uci.edu, markw@uci.edu
University of California, Irvine

Abstract: We present MAESF, a multi-agent essay scoring and feedback, as a teacher-centered AI tool for scoring and feedback. The system distributes evaluation across specialized agents and enables educators to interactively query and refine rubric-aligned assessments. Evaluated on the ASAP dataset, MAESF outperforms single-agent baselines and approaches human inter-rater reliability. By combining transparent scoring with teacher-in-the-loop interaction, MAESF reframes automated assessment as a collaborative process that augments teacher capacity and supports more consistent, timely feedback.

Introduction

Essay writing is central to developing students' critical thinking and communication skills, yet providing high-quality, individualized feedback remains one of the most time-consuming challenges teachers face. Scoring essays requires interpreting complex, multidimensional rubrics, a task prone to inconsistency even among trained raters (Kayapinar, 2014; Huang & Whipple, 2023). In large classrooms, these demands make it difficult to deliver timely formative feedback that supports student revision and growth (Hattie & Timperley, 2007; Nicol & Macfarlane-Dick, 2006). Recent advances in large language models (LLMs) have opened new possibilities for automating essay assessment (Li & Liu, 2024). Yet current single-agent approaches often evaluate writing superficially, lack transparency in their reasoning, and generate feedback with limited instructional value (Fallah et al., 2024). Critically, they offer teachers little control over the assessment process, functioning as black-box tools that sideline educator judgment rather than supporting it. We introduce MAESF, a multi-agent system that reimagines AES as a teacher-facing collaborative tool. Rather than replacing human judgment, MAESF is designed to augment teacher capacity by providing structured, rubric-aligned evaluations with transparent rationales and enabling educators to interact conversationally with the system to refine and validate outputs.

Method and data

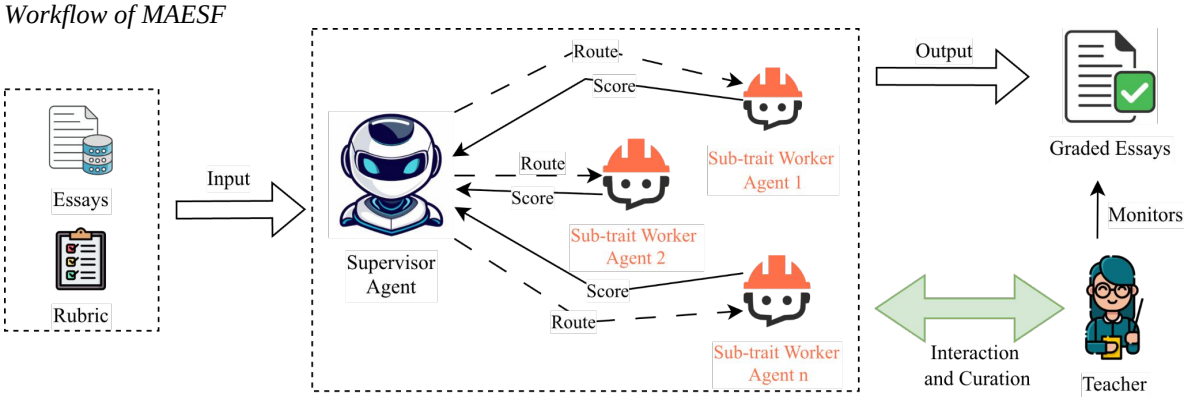
The MAESF framework organizes essay evaluation as a multi-agent assessment pipeline that decomposes scoring into interpretable, rubric-aligned components. To evaluate MAESF, we used a portion of Essay Set 8 from the Automated Student Assessment Prize (ASAP) dataset (Crossley et al., 2025), originally consisting of 723 Grade 10 essays written in response to a narrative prompt. Each essay was independently scored by two human raters across six writing traits—Ideas and Content, Organization, Voice, Word Choice, Sentence Fluency, and Conventions—on a 1-6 scale. A holistic score is derived from these trait-level ratings. This multi-trait, multi-rater structure enables fine-grained evaluation of both analytic and holistic scoring performance, as well as model–human agreement.

MAESF implements a hierarchical multi-agent architecture composed of a Supervisor Agent and multiple specialized Worker Agents. Given a student essay and scoring rubric, the Supervisor Agent first parses the rubric and decomposes it into its constituent traits. Each trait is then assigned to a dedicated Worker Agent, which operates independently to evaluate a single dimension of writing. This decomposition reduces cross-dimensional interference and allows each agent to focus on specific rubric criteria. Each Worker Agent produces three outputs: (1) a numeric sub-trait score aligned with rubric definitions, (2) an interpretable rationale explaining the scoring decision, and (3) targeted, actionable feedback to support student revision. These outputs are generated in parallel across agents. The Supervisor Agent subsequently aggregates them into a structured evaluation report that includes analytic scores, dimension-level explanations, and a synthesized holistic assessment.

A central feature of MAESF is its teacher-in-the-loop interaction design. After the report is generated, teachers engage with the system through a conversational interface (Figure 1). Rather than passively accepting results, educators can request clarification of specific scores, examine underlying rationales, flag inconsistencies, or trigger re-evaluation of selected traits. Queries related to individual dimensions are routed to the corresponding Worker Agent, while the Supervisor Agent handles holistic or cross-trait inquiries. This interaction positions MAESF as a teacher-centered assessment tool, enabling educators to actively curate, validate, and refine AI-generated evaluations before feedback is shared with students.

 © ISLS

Figure 1
Workflow of MAESF



Results

We evaluated MAESF against a conventional single-agent AES baseline using two backbone LLMs: GPT-4o-mini and GPT-5-mini. Model–human agreement on holistic scores was measured using quadratic weighted kappa (QWK), ±1 agreement, and Spearman's ρ, with the human–human agreement reported as a benchmark (Table 1). Across both backbones, MAESF consistently achieved stronger alignment with human raters than the baseline. With GPT-4o-mini, QWK improved from 0.367 to 0.521 against Rater 1 and from 0.312 to 0.449 against Rater 2. ±1 agreement increased from 25.45% to 35.27% (Rater 1) and 26.42% to 32.37% (Rater 2), while Spearman's ρ rose from 0.523 to 0.568 (Rater 1) and 0.471 to 0.505 (Rater 2). Even larger gains were observed with GPT-5-mini, where QWK climbed from 0.447 to 0.615 (Rater 1) and 0.378 to 0.548 (Rater 2); ±1 agreement from 27.25% to 44.95% and 28.49% to 39.42%; and ρ from 0.569 to 0.618 and 0.522 to 0.560, respectively. The strongest configuration, MAESF with GPT-5-mini, approached human inter-rater reliability on all three metrics, with QWK = 0.615 versus the human benchmark of 0.624 (Δ = 0.009), ±1 agreement of 44.95% versus 48.13% (Δ = 3.18 pp), and ρ = 0.618 versus 0.626 (Δ = 0.0077).

Table 1
Model vs. Human Agreement on Final (Holistic) Scores

Model	Method	QWK		±1 Agreement		Spearman ρ	
		Rater 1	Rater 2	Rater 1	Rater 2	Rater 1	Rater 2
GPT-4o-mini	SA	0.367	0.312	25.450	26.418	0.523	0.471
	MAESF	0.521	0.449	35.270	32.365	0.568	0.505
GPT-5-mini	SA	0.447	0.379	27.248	28.492	0.569	0.522
	MAESF	**0.615**	**0.548**	**44.952**	**39.419**	**0.618**	**0.560**
Human Agreement		0.624		48.133		0.626	

SA: Single Agent; IC: Ideas and Content; Org: Organization; Voc: Voice; WC: Word Choice; SF: Sentence Fluency; Conv: Conventions.

To contextualize these results, we compared MAESF with prior AES approaches using average QWK across raters. As shown, MAESF achieves an average QWK of 0.582, substantially outperforming prior single-agent methods such as Mansour et al., 2024 (QWK = ChatGPT: 0.313, Llama: 0.297). While supervised models such as Jiang et al., 2022 (QWK = 0.677) and Xie et al., 2022 (QWK = 0.779) report higher performance, these approaches rely on task-specific training. In contrast, MAESF operates in a zero-shot setting without fine-tuning, showing its advantage as a flexible and practical tool for real-world classroom use.

Discussions and conclusions

The results show that MAESF substantially improves alignment with human raters compared to single-agent approaches, approaching human-level agreement without task-specific training. Consistent gains across models and raters indicate that the multi-agent design provides a robust and generalizable approach to holistic essay scoring. Beyond performance, MAESF functions as a teacher-centered assessment tool. By generating rubric-aligned scores with interpretable rationales, it makes evaluation more transparent and supports teacher judgment. The conversational interface further enables educators to query and refine outputs, reframing automated scoring as a collaborative human–AI process rather than a black-box prediction. While MAESF demonstrates strong performance, automated scoring remains imperfect, and teacher oversight is essential.

 © ISLS

References

Crossley, S. A., Baffour, P., Burleigh, L., & King, J. (2025). A large-scale corpus for assessing source-based writing quality: ASAP 2.0. Assessing Writing, 65, 100954. doi:10.1016/j.asw.2025.100954

Fallah, A., Keramati, A., Nazari, M. A., & Fatemeh Sadat Mirfazeli. (2024). Automating Theory of Mind Assessment with a LLaMA-3-Powered Chatbot: Enhancing Faux Pas Detection in Autism. 365–372. https://doi.org/10.1109/iccke65377.2024.10874775

Hattie, J., & Timperley, H. (2007). The Power of Feedback. Review of Educational Research, 77(1), 81–112. https://doi.org/10.3102/003465430298487

Huang, J., & Whipple, P. B. (2023). Rater variability and reliability of constructed response questions in New York state high-stakes tests of English language arts and mathematics: implications for educational assessment policy. Humanities and Social Sciences Communications, 10(1). https://doi.org/10.1057/s41599-023-02385-4

Jiang, Z., Gao, T., Yin, Y., Liu, M., Yu, H., Cheng, Z., & Gu, Q. (2023, July). Improving Domain Generalization for Prompt-Aware Essay Scoring via Disentangled Representation Learning. In A. Rogers, J. Boyd-Graber, & N. Okazaki (Eds), Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 12456–12470). doi:10.18653/v1/2023.acl-long.696

Kayapinar, U. (2014). Measuring Essay Assessment: Intra-rater and Inter-rater Reliability. Eurasian Journal of Educational Research, 14(57). https://doi.org/10.14689/ejer.2014.57.2

Li, W., & Liu, H. (2024). Applying large language models for automated essay scoring for non-native Japanese. Humanities & Social Sciences Communications, 11(1). https://doi.org/10.1057/s41599-024-03209-9

Mansour, W. A., Albatarni, S., Eltanbouly, S., & Elsayed, T. (2024, May). Can Large Language Models Automatically Score Proficiency of Written Essays? In N. Calzolari, M.-Y. Kan, V. Hoste, A. Lenci, S. Sakti, & N. Xue (Eds), Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024) (pp. 2777–2786). Retrieved from https://aclanthology.org/2024.lrec-main.247/

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative Assessment and Self-regulated learning: a Model and Seven Principles of Good Feedback Practice. Studies in Higher Education, 31(2), 199–218. https://doi.org/10.1080/03075070600572090

Xie, J., Cai, K., Kong, L., Zhou, J., & Qu, W. (2022, October). Automated Essay Scoring via Pairwise Contrastive Regression. In N. Calzolari, C.-R. Huang, H. Kim, J. Pustejovsky, L. Wanner, K.-S. Choi, … S.-H. Na (Eds), Proceedings of the 29th International Conference on Computational Linguistics (pp. 2724–2733). Retrieved from https://aclanthology.org/2022.coling-1.240/

Acknowledgments

This work is based upon work supported by the National Science Foundation under Grant No. 23152984.

International Society of the Learning Sciences

Fostering Preservice Teachers' Design Thinking by Integrating Generative AI and Analytics-Supported Reflective Assessment in a Collaborative Inquiry Environment

Yuqin Yang, Central China Normal University, China, yangyuqin@ccnu.edu.cn
Linbaiyu Chen, Central China Normal University, China, baiyu_cl_apply@163.com
Daner Sun, The Education University of Hong Kong, Hong Kong S.A.R., dsun@eduhk.hk
Wangda Zhu, The Hong Kong Polytechnic University, Hong Kong S.A.R., wangda.zhu@polyu.edu.hk
Shiyan Pang, Central China Normal University, China, pangsy@ccnu.edu.cn

Abstract: This study examined how integrating generative AI (GenAI) and analytics-supported reflective assessment (ASRA) within a Knowledge Building (KB) environment supports preservice teachers' (PSTs) design thinking (DT). Participants were 110 PSTs in a semester-long instructional design course. We implemented a design that combined collaborative inquiry, delayed GenAI use, and reflective analytics, positioning GenAI as a design partner. Data from a DT scale and instructional design artifacts were analyzed. Results showed gains in problem discovery and definition and a shift toward more iterative design practices. The study offers a learning sciences account of integrating GenAI and ASRA to support DT in teacher education.

Introduction

Fostering PSTs' DT is essential for addressing complex instructional challenges, yet remains difficult in teacher education. Although GenAI and ASRA (Yang et al., 2024, Yang et al., 2020) show promise, their integration within collaborative inquiry is underexplored. This study positions DT as the core learning outcome within a KB environment (Scardamalia & Bereiter, 2022), where it is enacted through instructional design. GenAI functions as a cognitive partner to support idea generation and problem reframing, while ASRA, supported by KBDeX (KBDeX; Oshima et al., 2012), visualizations and reflective prompts, facilitates monitoring, evaluation, and regulation of design processes. Together, they form an integrated approach that supports epistemic agency and iterative improvement. This study examines whether PSTs develop DT through the combined use of GenAI and ASRA in a KB collaborative inquiry environment. This study aimed to address the research question: Do PSTs develop their DT through the combined use of GenAI and ASRA in a KB collaborative inquiry environment?

Method and results

Participants and instructional design

This study was conducted at a top-tier teacher education university in central China, whose core institutional mission is to prepare qualified K–12 teachers. The participants were 110 third-year PSTs enrolled in a compulsory undergraduate course. All participants were English education majors, a discipline that is highly gender-skewed in the Chinese teacher education context. Guided by the four theoretically grounded principles—epistemic agency, collective responsibility for idea improvement, human–AI co-agency and co-design, and embedded and transformative assessment, we developed an integrated pedagogical design that coordinated KB inquiry, GenAI-supported co-design, and ASRA. The design comprises three interrelated perspectives: (1) a design-oriented collaborative inquiry cycle structured around the DDIPT model (Discover–Define–Ideate–Prototype–Test); (2) GenAI-supported co-design to expand, critique, and refine instructional ideas; and (3) ASRA, including KBDeX-based discourse visualizations and metacognitive scaffolds to guide monitoring and regulation.

Data sources, analysis and results

To examine changes in PSTs' DT, we used two complementary data sources: self-reported DT and successive versions of group-based instructional design artifacts.

Self-reported DT was measured using the 29-item DT Competency Scale (Cai & Yang, 2023). The questionnaire was administered online at the beginning and end of the course. Of the 110 participants, seven cases were excluded due to incomplete or inattentive responses, resulting in 103 matched cases. Reliability was high (overall Cronbach's $\alpha = 0.931$). After confirming normality, paired-sample t-tests and Cohen's d were used to assess pre–post changes. Table 1 presents the means (M) and standard deviations (SD) of PSTs' DT before and after the combined use of GenAI and ASRA, respectively, and the paired-sample *t*-test results. Table 1 shows a statistically significant improvement of PSTs' DT from pre-test (M = 5.15, SD = 0.60) to post-test (M = 5.25, SD

 © ISLS

= 0.58), $t(102) = 2.151$, $p = .034$, with a small effect size (Cohen's $d = 0.21$). This indicates that the observed change was modest in magnitude and should be interpreted as gradual developmental movement rather than a large shift in self-perceived competency. Overall, these results indicate incremental growth in PSTs' early-stage DT processes, particularly problem exploration and definition, while later phases such as ideation and testing may require longer-term scaffolding or additional design cycles to become salient in self-report measures.

Table 1

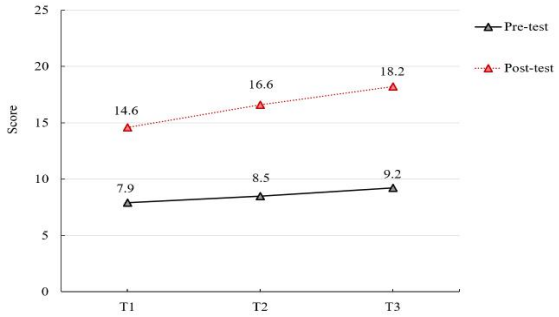
Changes in PSTs' Self-Reported Design Thinking (DT) Across the Intervention

	No. of Items	Pre-test		Post-test		t	df	p
		M	SD	M	SD			
Overall	29	5.15	0.60	5.25	0.58	2.151	102	0.034
Discover	6	5.25	0.46	5.37	0.63	2.258	102	0.026
Define	6	5.10	0.64	5.25	0.57	2.386	102	0.019
Ideate	6	5.13	0.68	5.21	0.71	1.371	102	0.174
Prototype	6	5.03	0.68	5.15	0.68	1.965	102	0.052
Test	5	5.24	0.67	5.27	0.61	0.547	102	0.586

To capture enacted DT, we analyzed lesson-plan design artifacts produced by 22 groups across three time points (T1, T2, T3). Each group developed three lesson plans over the semester. For each lesson, groups first completed a human-designed version and then revised it through human–AI co-design after human-only revision had plateaued. A rubric aligned with the DDIPT model and instructional design criteria was used to score all 132 lesson-plan versions, with inter-rater reliability above 92%. Figure 1 presents the PSTs' performance trends across the three time periods of lesson plan design (T1, T2, and T3). In the human–GenAI co-designed condition, mean scores increased steadily from 14.6 at T1 to 18.2 at T3. By contrast, in the human-designed condition, mean scores rose more slowly, from 7.9 to 9.2 over the same period. Because both design conditions were experienced by the same PST groups as part of a sequential design process, these trends reflect within-design developmental contrasts rather than differences between independent groups. This descriptive pattern suggests that the human–GenAI co-designed condition demonstrated stronger growth in instructional design performance, consequently, their DT competencies across the semester.

Figure 1

Developmental Trends Across Sequential Design Phases



Conclusions

This study found that PSTs' DT improved over time, with gains in problem discovery and definition and more integrated engagement across design phases. GenAI was most effective when used as a reflective partner rather than a solution provider, particularly when introduced after initial human design. ASRA supported monitoring and iterative refinement. Together, these findings underscore the importance of aligning GenAI use with reflective, collaborative pedagogical structures. This study extends learning sciences research by showing how GenAI and ASRA can be integrated within a KB environment to support DT development, advancing understanding of human–AI co-agency and reflective knowledge creation. Practically, it offers a principled approach for teacher education, demonstrating how GenAI–analytics–KB integration can foster reflective and collaborative design capacities. While further work is needed on scalability and transfer, it provides a foundation for designing AI-augmented, process-oriented learning environments.

 © ISLS

References

Oshima, J., Oshima, R., & Matsuzawa, Y. (2012). Knowledge Building Discourse Explorer: a social network analysis application for knowledge building discourse. *Educational technology research and development, 60*, 903-921. https://doi.org/10.1007/s11423-012-9265-2

Scardamalia, M., & Bereiter, C. (2022). Knowledge building and Knowledge creation. In R. K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (pp. 385–405). Cambridge University Press.

Yang, Y., Chan, C. K. K., Zhu, G., Tong, Y., & Sun, D. (2024). Reflective assessment using analytics and artifacts for scaffolding knowledge building competencies among undergraduate students. *International Journal of Computer-Supported Collaborative Learning*, 19, 231-272. https://doi.org/10.1007/s11412-024-09421-8.

Yang, Y., Du, Y., van Aalst, J., Sun, D., & Ouyang, F. (2020). Self-directed reflective assessment for collective empowerment among pre-service teachers. *British Journal of Educational Technology*, 51(6), 1960–1980.

 © ISLS

International Society of the Learning Sciences

Shared Epistemic Agency in A Collaborative Game for AI Literacy

Yue Shi, Lingfei Cao, Zhichun Liu
petra9@connect.hku.hk, lingfei.cao@connect.hku.hk, lukasliu@hku.hk
The University of Hong Kong

Abstract: This study examines how middle school students enacted shared epistemic agency in *Occultation*, a collaborative Turing-test game for AI literacy. Analyzing 353 chat messages from 28 students, we find that socio-emotional behaviors — trolling, spamming, and banter — directly triggered epistemic reasoning and regulative coordination, extending Damşa et al.'s (2010) framework. Rather than disruptions, these behaviors functioned as inquiry tools in a low-stakes, teacher-free environment. Findings suggest game-based contexts can productively leverage emotional engagement for collaborative AI literacy learning.

Introduction

As Generative Artificial Intelligence (GenAI) becomes more accessible and prevalent in education, the core challenge is not simply a knowledge deficit but a potential epistemic challenge. If students believe GenAI to be an objective and reliable knowledge source, they will lack the disposition to question its outputs, even when instructed to be critical (Fan et al., 2025). The goal of AI literacy should address *how to think* about it as a non-human knowledge agent through epistemic agency (Barzilai & Chinn, 2024; Zhang et al., 2022).

Well-designed educational games can function as high-agency, low-stakes environments that are uniquely suited to activating shared epistemic agency (Gee, 2003). Unlike teacher-directed settings where epistemic goals are externally defined, games position learners to define, test, and revise their own assumptions. We designed *Occultation*, a collaborative Turing-test game that positions AI as an object of collective inquiry, and ask: *How do learners enact shared epistemic agency in a collaborative game for AI literacy?*

Theoretical perspective

To understand how learners collaboratively develop ways of thinking about AI, we ground our work in the concept of shared epistemic agency. We chose this framework over individual epistemic agency approaches (Zhang et al., 2022) because our game context proposes collective collaboration that no single player can identify the hidden AI alone. Unlike knowledge-building frameworks (Scardamalia & Bereiter, 2006) that focus on artifact production, Damşa et al. centers the process of agentic inquiry – which fits a game where learning happens through live conversation.

This agency manifests through two dimensions: (a) the epistemic dimension, focused on the collaborative creation of knowledge, and (b) the regulative dimension, focused on managing the group's process for producing that knowledge (Damşa et al., 2010). Yet many students' behaviors, such as emotional reactions, banter, trolling, did not fit either dimension. The regulative dimension captures how the group organizes its process (e.g., turn-taking, voting, misdirection), whereas the behaviors we observed generated relational signals — emotional evidence that peers used as epistemic data (e.g., "why would AI curse?"). This functional distinction, not merely categorical difference, justifies a separate dimension. Following Yang & Markauskaite (2023), who argue that emotions are central to epistemic agency, not just background noise, we identified a third socio-emotional dimension. We treat this as an exploratory extension that needs further study.

Method

The game and data collection

Occultation is a multiplayer conversational Turing-test-like game where players distinguish between human survivors and AI infiltrators. The game assigns two roles: Interrogators (identify the AI) and Traitors (protect the AI by misleading others). Twenty-eight students (aged 12-13) from a summer program in a suburban area of eastern China participated. Students were placed into six online chat rooms of four to five students. We focused on the conversational log that happened in one game session of a total of 353 messages.

Data analysis

We conducted an inductive-deductive analytical approach (Fereday & Muir-Cochrane, 2006) to analyze 353 chat messages from six groups. Two coders independently open-coded the transcript on message level to capture

 © ISLS

emergent behaviors and generated an initial codebook that included reasoning, questioning, and trolling. We then mapped codes to Damşa et al.'s (2010) two dimensions - epistemic Dimension (knowledge-related), and Regulative Dimension (process-related). Two coders independently coded all 353 messages, achieving a percent agreement of 89.36%, with disagreements resolved through discussion. However, a substantial portion of behaviors did not align with either dimension. Many behaviors did not fit either dimension, including emotional reactions, banter, and trolling produced their own agentic actions that shaped both knowledge work and group process. Following Yang & Markauskaite (2023), we identified these as a third socio-emotional dimension.

Findings

Students enacted shared epistemic agency through three dimensions — epistemic, regulative, and socio-emotional — that triggered one another rather than running in parallel. Table 1 summarizes the coding scheme.

Table 1

Shared Epistemic Agency in a Collaborative Turing-Test Game

Dimension	Description	Example
Epistemic (n = 88, 24%)	Knowledge-related actions — learners introduce, test, and evaluate ideas about AI identity	"Banana Captain replies too fast — must be the AI"; "Did you finish your homework?"
Regulative (n = 165, 45%)	Process-related actions — learners steer, organize, and monitor the group's collective inquiry	"1, 2, 3 — everyone answers in order"; "I am AI or Traitor"
Socio-emotional (n = 113, 31%)	Relation-related actions — emotional and disruptive behaviors that generate signals constitutive of inquiry	"Start the excavator" × 4 times; "Fk your excavator"; "Wait, why would AI curse?"

Epistemic dimension actions showed students taking ownership of knowledge work rather than passively guessing (Damşa et al., 2010). Students hypothesized that "quick and sloppy" replies indicated nonhuman authorship, then tested this by asking "Did you finish your homework?" — drawing on shared school knowledge to probe authenticity.

Regulative dimension actions showed students, not teachers, initiating productive struggle (Chen, 2022). One student declared "I am AI or Traitor" to sow deliberate confusion, testing whether peers could hold their reasoning under pressure, while others typed "1,2,3..." to set speaking order and proposed "Let's vote directly later."

Socio-emotional dimension made up 31% of all messages and directly shaped inquiry. This pattern appeared across all six groups, albeit with varying intensity. "What the heck?!" drew peers' attention to re-examine their criteria. Spamming ("Start the excavator," ×4 times) triggered "Fk your excavator," which sparked: "Wait, why would AI curse?" Trolling produced visible anger that students read as evidence of human identity — a form of applied AI literacy, as students used emotional provocation as a live Turing-test probe to distinguish human from AI behavior. These behaviors emerged organically in a low-stakes game; we do not advocate their deliberate use without careful scaffolding.

The three dimensions triggered one another. One student declared "I am AI or Traitor" — peers responded not with a direct question but with banter ("No hard feelings, just kidding?") to probe the claim's authenticity. That exchange then prompted a call to vote, showing how regulative strategy enabled socio-emotional probing, which in turn informed collective regulation.

Discussion

Frustration and playfulness directly triggered epistemic reasoning and regulative coordination, extending Damşa et al.'s (2010) framework and aligning with work on epistemic emotions in co-inquiry (Candiotto, 2022). Occultation made this possible by removing the teacher and the pressure for one correct answer — students need to experience epistemic uncertainty about AI, not just be told to think critically. For educators, socio-emotional disruption can be a resource, not just a problem to manage. Findings are exploratory — one session, 28 students, no interviews. Future work should test these patterns across game designs, examine how Traitor and Interrogator roles shape different epistemic strategies, and include student interviews to validate interpretive claims.

References

Barzilai, S., & Chinn, C. A. (2024). The AIR and Apt-AIR Frameworks of Epistemic Performance and Growth: Reflections on Educational Theory Development. *Educational Psychology Review*, *36*(3), Article 91. https://doi.org/10.1007/s10648-024-09927-5

Candiotto, L. (2022). Epistemic emotions and co-inquiry: A situated approach. *Topoi*, *41*(5), 839-848.https://doi.org/10.1007/s11245-021-09789-4.

Chen, Y. (2022). Epistemic uncertainty and the support of productive struggle during scientific modeling for knowledge co-development. *Journal of Research in Science Teaching*, *59*(3), 383–422. https://doi.org/10.1002/tea.2173

Damşa, C. I., Kirschner, P. A., Andriessen, J. E., Erkens, G., & Sins, P. H. (2010). Shared epistemic agency: An empirical study of an emergent construct. *The Journal of the Learning Sciences*, *19*(2), 143-186.https://doi.org/10.1080/10508401003708381

Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., & Gašević, D. (2025). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology*, *56*(2), 489-530. https://doi.org/10.1111/bjet.13544

Fereday, J., & Muir-Cochrane, E. (2006). Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *International Journal of Qualitative Methods*, *5*(1), 80–92. https://doi.org/10.1177/160940690600500107

Gee, J. P. (2003). What video games have to teach us about learning and literacy. *Computers in Entertainment CIE*, *1*(1), 20–20. https://doi.org/10.1145/950566.950595

Scardamalia, M., & Bereiter, C. (2005). Knowledge Building: Theory, Pedagogy, and Technology. In R. K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (pp. 97–116). Cambridge University Press. https://doi.org/10.1017/CBO9780511816833.008

Yang, H., & Markauskaite, L. (2023). Preservice teachers' perezhivanie and epistemic agency during the practicum. *Pedagogy, Culture & Society*, *31*(4), 647–668. https://doi.org/10.1080/14681366.2021.1946841

Zhang, J., Tian, Y., Yuan, G., & Tao, D. (2022). Epistemic agency for costructuring expansive knowledge-building practices. *Science Education*, *106*(4), 890–923. https://doi.org/10.1002/sce.21717

International Society of the Learning Sciences

Exploring Tangible Interfaces for Learning 3D Geometry and Basic Structural Engineering Concepts

Arooj Zaidi, Graduate School of Keio Media Design, arooj.zaidi@keio.jp
Sandra Yu Okita, Teachers College, Columbia University, so2269@tc.columbia.edu
Giulia Barbareschi, University of Duisburg-Essen, giulia.barbareschi@uni-due.de
Chihiro Sato, Graduate School of Keio Media Design, chihiro@kmd.keio.ac.jp
Junichi Yamaoka, Graduate School of Keio Media Design, yamaoka@kmd.keio.ac.jp

Abstract: This paper explores ways tangible interfaces can be used as a scaffolding tool for children to better understand 3D geometrical concepts during familiar cube structural activities, and bridge building activities. The study found that through embedded visual and audio feedback, six- to twelve-year-old children were able to identify and address gaps in their own learning, self-correct, and increase their understanding of 3D geometric concepts. Children were also able to successfully apply their understanding to basic engineering concepts in structural building activities. Overall, this study helped to identify principles for designing a framework for tangible interfaces able enhance children's learning experiences of 3D geometry.

Introduction and background

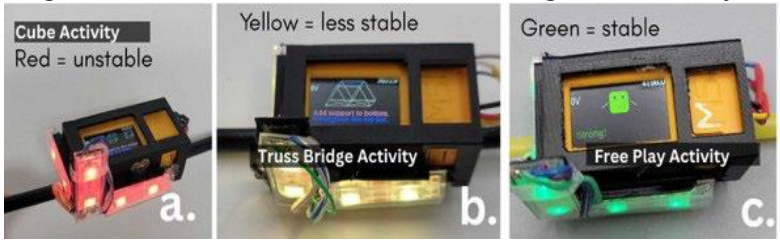
Understanding geometric concepts in two and three dimensions can be challenging for children, as traditional representations often fail to effectively convey spatial relationships (Battista, 2007; Sarama & Clements, 2009). While hands-on manipulatives like geoboards make abstract ideas more tangible, they usually lack built-in scaffolding for independent learning (Resnick, 1998; Manches & Price, 2011). Tangible User Interfaces (TUIs) enhance these tools by incorporating computational feedback, enabling learners to explore concepts through embodied interaction with real-time guidance (Schneider & Blikstein, 2015). However, designing TUIs that connect learning concepts with familiar everyday interactions and real-world applications (e.g., structural engineering) remains a challenge. This gap results in missed opportunities for children to develop independent 2D and 3D geometry skills, emphasizing the need for innovative approaches that integrate hands-on exploration with support for self-directed learning. Drawing on constructivism, constructionism, sociocultural theory, and embodied cognition, this study examines how tangible interfaces can enhance children's spatial reasoning and self-monitoring while building stable 3D structures. It addresses two questions: (RQ1) How can tangible interfaces with real-time feedback support children's self-monitoring and understanding of structural stability? and (RQ2) What design principles foster self-monitoring and self-correction in tangible geometry learning environments?

TangiBuild (3D Geometry)

Extending our earlier work on TIEboard (Zaidi et al., 2025), a TUI leveraging familiar geoboard-based interactions for 2D geometry, TangiBuild uses Strawbees (Strawbees, 2025) construction materials to support reasoning about structural stability in hands-on 3D construction. A clip-on module built with the M5StickC Plus2 uses an IMU sensor to detect tilt and provides real-time multimodal feedback through a traffic-light LED system, haptic vibration, and short onscreen hints. Rather than giving exact answers, the system offers partial guidance that encourages children to identify weak points and test possible reinforcements themselves. Children used TangiBuild across three activities of increasing complexity: cube construction, bridge building, and a free-play tower challenge.

Figure 1

TangiBuild Feedback and Hints Across Cube, Bridge, and Free-Play.



 © ISLS

Study

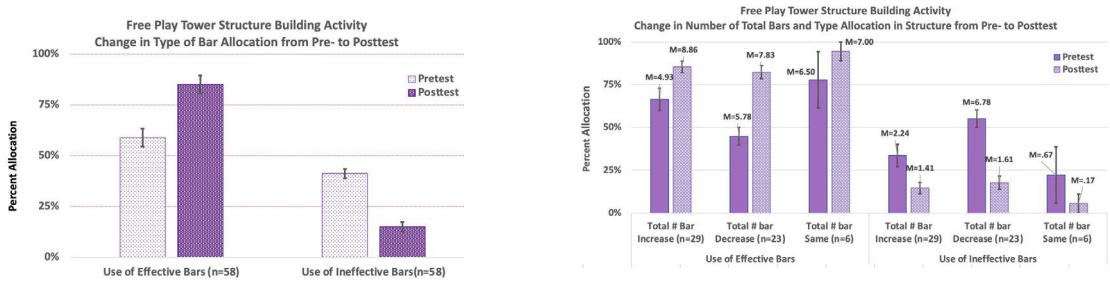
Fifty-eight children between 6-12 years of age participated in an after-school program at an urban elementary school in Chiba, Japan. Across six sessions, children engaged in cube, bridge, and free-play tower building activities. Each activity started (pre-test) and concluded (post-test) with a brief 4 item question for the children. For the cube and bridge activities, the surveys required children to rank structures from weakest to strongest. Children needed to examine how the bars were placed in the cube or bridge (e.g., horizontally or creating the shape of a triangle) to determine if it is reinforcing the structure to be strong or weak. During the free play tower activity, children were given a preconstructed tower and asked to reinforce the structures by placing the bars. We tracked the number of diagonal and horizontal bars they added before (pre-test) and after the activity(post-test). Bars that were placed diagonally were counted as effective bar placement, and bars that were placed horizontally were counted as ineffective placement. Video recordings and field observations captured how children interpreted feedback, negotiated design choices, and applied reinforcement strategies.

Findings

Children demonstrated improved understanding of structural stability across activities of increasing complexity. The average percent accuracy score involving children's cube activity(n=49) from pretest (M=68%, SD=.395) to the posttest (M=98%, SD=.099) significantly increased at $t(48) = -5.392$., $p < 0.001$. The average percent accuracy score for the bridge activity from pretest (M =78%, SD=0.36) to the posttest (M= 99%, SD= 0.069) significantly increased at $t(52) = -4.285$, $p < 0.001$. In the free play tower activity (n = 58), we assessed children's ability to reinforce a pre-constructed tower with effective bar placements. Each child's effective and ineffective placements were individually calculated. A paired-sample t-test revealed significant learning gains: effective placements increased from an average of 58% (M=58%, SD=0.340) in the pre-test to 85% (M=85%, SD=0.175) in the posttest ($t(57) = -7.779$, $p < 0.001$), while ineffective placements decreased from 40% (M=40%, SD=0.335) to 15% (M=15%, SD=0.175) ($t(57) = 7.741$, $p < 0.001$). This indicated that children learned to implement effective strategies (See Figure 2a). Among the participants, 29 children increased their total placements from an average of 7 bars pre-test to 10 post-test, improving their effective placements while reducing ineffective ones. In contrast, 23 children who reduced their placements from 13 to 9 also increased their effective bars, suggesting selfcorrection. Even those maintaining the same number (n=6, average of 7) might have self-corrected (See Figure 2a). Observations further showed that children interpreted feedback collaboratively and used it to test alternative solutions. For example, when a child questioned why the lights weren't turning green, peers suggested adding diagonal support and testing a new placement. After this revision, the child realized the structure had stabilized. This illustrates that hands-on manipulation, along with visual and sound feedback, helps students address learning gaps, self-correct, and deepen their understanding of 3D geometric concepts.

Figure 2

Change in Number of Total Bars From Pre- to Post- Test and See the Type of Bar Placement Made In Free Play.



(a) (b)

Design implications

Findings outline key considerations for designing tangible learning systems that enhance self-monitoring in construction activities. Feedback should encourage revision, as seen in TangiBuild where visual cues prompted children to rethink and adjust their structures. Using familiar materials like Strawbees reduced barriers for learners. Open-ended activities helped transition from guided exploration to independent strategy use, while tangible feedback environments fostered collaborative reasoning. These insights emphasize the role of feedback driven interaction in developing strategic reasoning for 3D structural concepts, with future research needed in diverse classroom settings.

 © ISLS

References

Battista, M. T. (2007). The development of geometric and spatial thinking. In F. Lester (Ed.), Second handbook of research on mathematics teaching and learning (pp. 843–908). Charlotte, NC: Information Age Publishing.

Manches, A., & Price, S. (2011). Designing learning representations around physical manipulation: Hands and objects. In Proceedings of the 10th International Conference on Interaction Design and Children (pp. 81–89). New York, NY: ACM. https://doi.org/10.1145/1999030.1999042

Papert, S., & Harel, I. (1991). Situating constructionism. constructionism, 36(2), 1-11.

Resnick, M. (1998). Technologies for lifelong kindergarten. Educational Technology Research and Development, 46(4), 43–55. https://doi.org/10.1007/BF02299648

Sarama, J., & Clements, D. H. (2009). Early childhood mathematics education research: Learning trajectories for young children. New York, NY: Routledge.

Schneider, B., & Blikstein, P. (2015). Flipping the flipped classroom: Developing a design framework for tangible computing in mathematics. Journal of the Learning Sciences, 24(1), 10–46. https://doi.org/10.1080/10508406.2014.964019

Strawbees. (n.d.). Strawbees building and robotics kits. Retrieved June 2025, from https://strawbees.com

Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. Cambridge, MA: Harvard University Press.

Zaidi, A., Barbareschi, G., Kunze, K., Pai, Y. S., & Yamaoka, J. (2025). TIEboard: A Digital Educational Tool for Kids Geometric Learning. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 9(2), 1-27. https://doi.org/10.1145/3729478

 © ISLS

International Society of the Learning Sciences

Developing Postgraduates' Feedback Literacy through Multiple-sourced Feedback Practices in an L2 Writing Class

Chunlin LEI, Fuxin KONG
leichunlin@suibe.edu.cn, 2835225514@qq.com
Shanghai University of International Business and Economics

Abstract: This study explored a multi-sourced (peer, AI, teacher) feedback model in enhancing student feedback literacy. Participants were 18 postgraduates registered in an L2 business English writing course. Findings revealed the complementary roles of different sources: peer review developed evaluative judgment, AI addressed linguistic accuracy, and teacher feedback targeted content and logic. The model extends beyond single or doubled sourced feedback practice and has pedagogical implications on instructional design for L2 writing classes.

Introduction

As a core component of second language (L2) writing instruction, feedback has consistently remained a central topic of interest for researchers (Carless & Boud, 2018; Hyland & Hyland, 2019; Yu et al., 2023). Scholars argue that by developing feedback literacy, students can better understand the value of feedback, make informed judgments, regulate affective responses and take actions, thereby enhancing their ability to use feedback for improvement (Han & Xu, 2020; Molloy et al., 2020). Despite the growing recognition of its significance, developing student feedback literacy in L2 writing remains at an exploratory stage and empirical evidence is still limited, especially in the Era of AI.

Traditional English writing instruction often relies on teacher feedback, which influences student feedback literacy through cognitive, behavioral, and affective dimensions. With a paradigm shift to student-centered learning, more peer feedback practices are implemented in writing instruction. By evaluating peers' writing, students not only deepen their understanding of writing standards, but also improve their ability to critically appreciate feedback, self-regulate and manage their own feedback processes (Zhang & Gao, 2025). More recent research has shown that AI-generated feedback has emerged as a strong plus to the sources of feedback. Compared to teacher or peer feedback, AI feedback offers students more accurate information about student learning performance (Lu et al., 2023) and enhance learners' acceptance of feedback (Banihashem et al., 2024).

However, empirical research on the role of AI-generated feedback in developing students' feedback literacy remains limited. In particular, few studies have systematically explored how teacher, peer, and AI feedback can be integrated to form a coherent multi-sourced feedback system that enhances students' feedback literacy. To address this gap, the present study investigated a tri-sourced feedback within a structured instructional framework, aiming to reveal how their complementary strengths contribute to the development of students' interpretive, applicative, and affective feedback capacities. This integrative approach not only contributes to a more nuanced understanding of feedback literacy development but also carries significant pedagogical implications for the integrated and effective use of multi-sourced feedback in L2 writing classrooms.

Research design

This study mainly focused on how the multi-sourced feedback in English writing promotes learners' feedback literacy and improvement on writing. Specifically, it included two research questions: (1) What are the characteristics of peer, AI and teacher feedback in postgraduate English writing class? (2) How does the multi-sourced feedback practices (peer, AI and teacher) improve postgraduates' feedback literacy?

The participants of the study were 18 Year-1 postgraduates majored in applied linguistics at a Chinese Mainland university. In a course titled "Research and Practice of Business English Writing", multi-sourced feedback practices were designed and implemented throughout the course. Specifically, students were systematically engaged in a series of feedback activities involving giving, receiving, interpreting, and acting upon feedback from multiple sources, including peers, AI tools (e.g., ChatGPT), and the instructor. For peer feedback, a technological platform called Peerceptiv was used to facilitate the peer feedback giving and receiving processes. For AI feedback, ChatGPT was utilized to generate comments for students' writing improvement. The teacher's feedback was conducted after peers' and AI's feedback. The interfaces of feedback process involving peers (Peerceptiv), AI (ChatGPT) and teacher (Revision mode of Word) are shown in Figures 1, 2 & 3 respectively. Students were also required to write reflective journals when they submitted each of the revised version. In

 © ISLS

reflective journals, students articulated their experiences, feelings, understanding and actions interacting with different types of feedback and compared the merits and demerits of different types of feedback in their writing.

Throughout the course, three tasks were designed, following a circle of "assigning the writing task → students submitting Draft 1→conducting peer review by Peerceptiv→ composing revised Draft 2 and reflective journal → AI generated feedback to the revised Draft 2→ composing further revised Draft 3 and reflective journal → teacher's feedback → finishing final Draft 4 and reflection". Each task/circle lasted four weeks. The data included records of feedback activities, drafts of writing, and reflective journals at various stages.

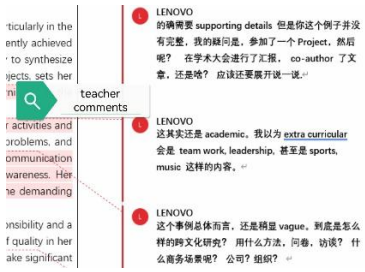
Figure 1
The Peerceptiv Platform



Figure 2
AI Feedback

Original (email)	AI-improved version
"I received a letter yesterday..."	"I am writing to respectfully request..."
"I really appreciate all your help..."	"I deeply appreciate the opportunities..."
"my request is that I could still enjoy..."	"I kindly propose the following arrangement..."
"but I am really not prepared for the leaving so soon"	"allow me to integrate fieldwork findings..."

Figure 3
Teacher Feedback



Preliminary findings

This poster only summarized a few themes pertaining to student feedback literacy in the dataset.

(1) Characteristics of peer, AI and teacher feedback. Peer feedback, conducted via Peerceptiv, was highly structured and accessible due to the provided rubrics. Students found it manageable to score and evaluate based on criteria. However, the tendency to rely on the rubric also limited its depth. Most students offered minimal written comments, and those that were provided were often general, focusing on surface-level issues like grammar. In contrast, AI-generated feedback was comprehensive at the linguistic level. It produced a fully revised version of the text, primarily offering substitutions for vocabulary and sentence structures. It paid little attention to higher-order concerns such as the appropriateness of content and argumentative coherence. The teacher's feedback shifted decisively to content and logic, as compared to the previous surface errors such as grammar. For example, in the recommendation letter writing, the teacher pinpointed inadequacies in content, logical flaws, and structural weaknesses. This strategic focus guided students toward a deeper understanding of the writing task.

(2) Role of peers, AI and teacher feedback in promoting student feedback literacy. The integration of multi-sourced feedback played a distinct and complementary role in developing postgraduates' feedback literacy. The act of providing peer feedback within the structured Peerceptiv system was crucial for helping students make sense of assessment criteria, thereby building their evaluative judgment. Repeatedly managing affective responses to diverse peer comments fostered greater emotional resilience and a more objective view of critique. AI feedback empowered students in the practical use of feedback for immediate linguistic improvement, encouraging proactive engagement with tools to refine their writing. Most significantly, the teacher's role evolved to focus on content and logic, challenging students to make judgements on higher-order concerns. This guided them beyond surface-level corrections, cultivating the critical discernment needed to take action for substantive, meaning-level revisions, which is the cornerstone of advanced feedback literacy.

Implications

The findings of this study carry significant implications for L2 writing pedagogy and curriculum design. They underscore that an integrated, multi-sourced feedback system is not merely beneficial but essential for developing comprehensive feedback literacy. The synergy between peer, AI, and teacher feedback creates a scaffolded learning ecosystem where each source addresses distinct aspects of the writing process. Peers foster evaluative judgment and resilience, AI provides immediate, low-stakes linguistic support and fosters learner autonomy, and the teacher guides high-order conceptual and rhetorical development.

For educators, this model necessitates a strategic redesign of writing tasks and feedback practices. It is crucial to leverage technological tools like Peerceptiv and AI to streamline the feedback process and manage teacher workload, thereby freeing up instructor capacity for more profound, content-focused interventions. However, this integration must be implemented with clear guidance. Students require training to use AI critically and avoid over-reliance, while teachers must refine their role to provide the unique, insightful commentary that machines cannot. Ultimately, a deliberately balanced feedback ecology is key to preparing students for the complex literacy demands of the future.

 © ISLS

References

Banihashem, S. K., et al. (2024). Feedback sources in essay writing: peer-generated or ai-generated feedback?. *International Journal of Educational Technology in Higher Education, 21*.

Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, 43(8), 1315–1325.

Han, Y., & Xu, Y. (2020). The development of student feedback literacy: The influences of teacher feedback on peer feedback. *Assessment & Evaluation in Higher Education*, 45(5), 680–696.

Hyland, K., & Hyland, F. (Eds.). (2019). Feedback in second language writing: Contexts and issues. Cambridge University Press.

Lu, X., et al. (2023). AI-Assisted Feedback for Writing: Accuracy and Effectiveness. *Educational Technology & Society*.

Molloy, E., Boud, D., & Henderson, M. (2020). Developing a learning-centered framework for feedback literacy. *Assessment & Evaluation in Higher Education*, 45(4), 527–540. https://doi.org/10.1080/02602938.2019.1667955

Yu, S., Yuan, K., & Wu, P. (2023). Revisiting the conceptualizations of feedback in second language writing: A metaphor analysis approach. *Journal of Second Language Writing*, 59, 100961.

Zhang, Y., & Gao, Y. (2024). Exploring the dynamics of student engagement with receiving peer feedback in L2 writing. *Assessing Writing*, 60(5), 100842

 © ISLS

International Society of the Learning Sciences

Introducing IDEAL: An Intelligent Design Environment for Active Learning

Chenxuan Meng, James D. Slotta
chenxuan.meng@mail.utoronto.ca, jim.slotta@utoronto.ca
University of Toronto

Abstract: Text-based curriculum tools inadequately represent spatial, temporal, and interactive dimensions of active learning design in active learning classrooms. Although generative AI offers accessible visualization options, existing approaches prioritize aesthetic plausibility over spatial correspondence with real classrooms. This poster presents the AI-assisted storyboarding component of IDEAL (Intelligent Design Environment for Active Learning), developed over three years of iterative design from 2022 to 2025. Our photobook-based approach constrains AI generation using systematic documentation of a specific classroom, including furniture inventory, technology layout, and spatial configurations, producing planning artifacts that help teachers judge the spatial plausibility of instructional scenarios before implementation. Exploratory validation with one experienced instructor suggests proof-of-concept feasibility, while also surfacing design tensions around fixation and cognitive load.

Background and design rationale

Equipped with flexible furniture and digital technologies, active learning classrooms (ALCs) support innovative pedagogical practices. Educators seek to design more interactive and contextualized learning experiences in these environments. Learning design research positions teachers as designers (Wasson & Kirschner, 2020); however, teachers' spatial competency remains a constraining factor in designing curriculum that can effectively leverage learning environments (Byers, 2021). Diagrammatic tools (e.g., LDVS, LAMS, Learning Designer) support alignment and collaboration (Agostinho, 2011; Conole & Wills, 2013; Dalziel et al., 2016), and storyboarding adds audience-facing narratives (Jacobs & Zmuda, 2023). But existing tools fail to capture spatial orchestration, for example how activities unfold across room configurations, technology placements, and student movement. Although generative AI opens new visualization possibilities, current applications prioritize inspiration over spatial reasoning.

We introduced Intelligent Design Environment for Active Learning (IDEAL), an emerging design support system that integrates spatial, technological, and pedagogical considerations for active learning curriculum design. This poster focuses on its AI-assisted storyboarding component, which addresses a crucial challenge: teachers designing for active learning need to anticipate whether a specific activity is feasible in a particular classroom, accounting for spatial capacity and constraints, circulation paths and furniture orientation. Generic text-to-image AI prompts produce aesthetically plausible but spatially unanchored outputs that vary arbitrarily across panels. The component therefore grounds AI generation in authentic classroom documentation, producing storyboards that support pedagogy-space fit reasoning. This poster explores two questions: *(1) How can AI support the creation of pictorial storyboards to represent spatial-temporal dimensions of active learning in support of teachers' design process? (2) How can such AI-generated storyboards support teachers in curriculum design?*

Iterative development: from 3D modeling to AI-assisted storyboarding

The IDEAL storyboarding component evolved through four phases from 2022, shifting from managing technical complexity toward integrating AI with architectural spatial documentation. Initial 3D modeling with Rhino (Figure 1) demonstrated the value of spatial-temporal visualization but required prohibitive specialized expertise. Generic AI generators (DALL-E 2, Midjourney) lacked spatial authenticity, producing spaces disconnected from real classrooms (Figure 2a). Photoshop AI improved background consistency using real classroom photo base, yet furniture misalignment persisted (Figure 2b). Combining Unreal Engine with DALL-E further improved spatial consistency but introduced inconsistent scale across panels (Figure 2c). Across all phases, the core problem was not tool selection but the absence of systematic spatial documentation. Phase 3 therefore introduced the Classroom Reference Library: seven photographs systematically documenting one active learning classroom, capturing all architectural features and constraints, furniture inventory, and technology infrastructure (Table 1). A five-stage workflow bridges teacher intentions and spatially grounded generation: capturing design intentions, identifying spatial decisions, constructing narrative sequences, matching panels to reference library images, and generating images via Gemini 2.5 Flash (Figure 3).

 © ISLS

Table 1

Classroom Reference Library Documentation

Category	Key Documentation
Spatial Parameters	18m × 9.6m (172m²), max capacity 60
Architectural Elements	Lime green north wall, neutral white other walls, two-toned flooring, acoustic ceiling with panel lights
Furniture Inventory	Verb flip-top tables (60"×22"), Tenor chairs (mobile), B-Free lounge sofas, Openest Chick poufs, mobile room dividers
Technology Infrastructure	Large digital display on portable stand, wall-mounted whiteboards, ceiling projector, integrated power outlets
Documented Configurations	Small group collaboration (tables in clusters), whole-class presentation (U-shape), gallery walk (wall space), individual work (lounge zone)

Generic text-to-image AI does not reliably preserve spatial proportions, furniture orientation, technology placement, or circulation constraints, instead focusing on aesthetic plausibility. Our photobook-grounded approach constrains generation through documented room-specific features, anchoring each panel to relevant Classroom Reference Library views (e.g., gallery walk panels matched to views with expansive whiteboard space), with prompts emphasizing spatial context: "Students walking around a classroom, discussing posters taped to a long white wall. Wide shot, academic exchange." The output is therefore not an imagined scene, but a planning artifact supporting pedagogy-space fit reasoning: teachers can judge whether an activity is feasible in this room, how students move, and whether spatial constraints undermine their pedagogical intent.

Figure 1

Prelude Storyboard

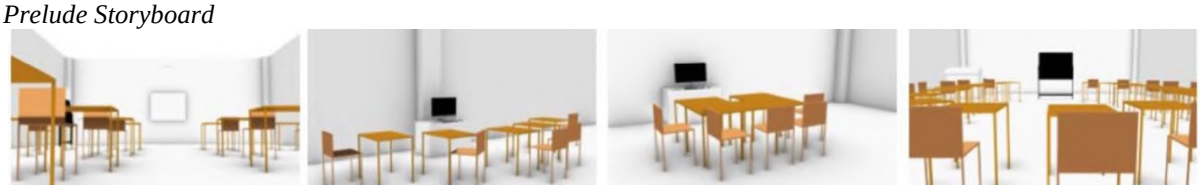
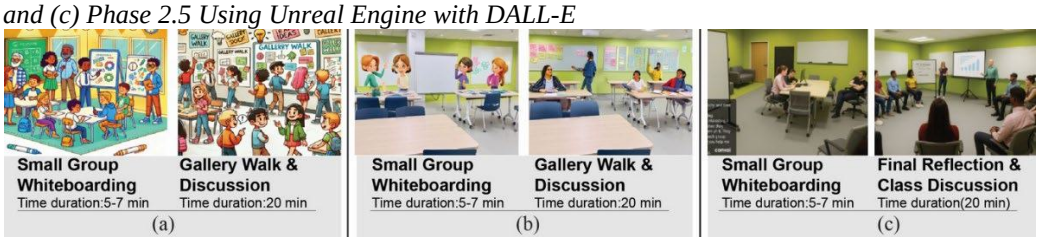


Figure 2

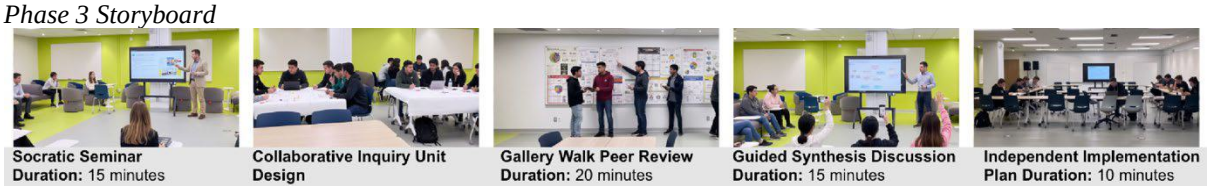
Evolution of AI-Assisted Storyboarding: (a) Phase 1 Using DALL-E 2, (b) Phase 2 Using Photoshop AI, and (c) Phase 2.5 Using Unreal Engine with DALL-E



Two representative panels from each phase are shown.

Figure 3

Phase 3 Storyboard



Exploratory validation and conclusion

To assess feasibility, we generated storyboards based on an experienced instructor's graduate course design. The instructor noted that the storyboard "made the flow between multi-screen displays and discussion zones concrete" and helped visualize "who does what when with which technology." While this single exploratory case cannot be generalized, it provides proof-of-concept that photobook-based contextualization addresses authentic design challenges in active learning environments. Design tensions remain: highly realistic outputs may encourage fixation over iterative exploration, and multi-stage workflows may increase cognitive load for novice designers. IDEAL is still in progress, validated in only one classroom with one instructor; transferability, long-term use, and scalability remain untested. Future work will pursue systematic comparisons, cross-context validation, and longitudinal studies linking storyboard-supported design to learning outcomes.

 © ISLS

References

Agostinho, S. (2011). The use of a visual learning design representation to support the design process of teaching in higher education. Australasian Journal of Educational Technology, 27(6), 961–978. https://doi.org/10.14742/ajet.923

Byers, T. (2021). What does teaching and learning look like in a variety of classroom spatial environments? In W. Imms & T. Kvan (Eds.), *Teacher transition into innovative learning environments: A global perspective* (pp. 187–201). Springer. https://doi.org/10.1007/978-981-15-7497-9_16

Conole, G., & Wills, S. (2013). Representing learning designs – making design explicit and shareable. Educational Media International, 50(1), 24–38. https://doi.org/10.1080/09523987.2013.777184

Dalziel, J. (2011). Visualising Learning Design in LAMS: A Historical View. Teaching English with Technology, 11(1), 19–34.

Dalziel, J., Conole, G., Wills, S., Walker, S., Bennett, S., Dobozy, E., Cameron, L., Badilescu-Buga, E., & Bower, M. (2016). The Larnaca Declaration on Learning Design. Journal of Interactive Media in Education, 2016(1). https://doi.org/10.5334/jime.407

Huizinga, T., Handelzalts, A., Nieveen, N., & Voogt, J. M. (2014). Teacher involvement in curriculum design: Need for support to enhance teachers' design expertise. Journal of Curriculum Studies, 46(1), 33–57. https://doi.org/10.1080/00220272.2013.834077

Jacobs, H. H., & Zmuda, A. (2023). Storyboarding your curriculum. Educational Leadership, 80(5), 22–27.

Wasson, B., & Kirschner, P. A. (2020). Learning Design: European Approaches. *TechTrends*, *64*(6), 815–827. https://doi.org/10.1007/s11528-020-00498-0

 © ISLS

International Society of the Learning Sciences

Analyzing Human–AI Interaction Patterns in Generative-AI-Supported Inquiry-Based Learning

Ayano Ohsaki, Tohoku University, aohsaki@ohsaki-lab.net
Hideya Matsukawa, Tohoku University, hideya.matsukawa.a5@tohoku.ac.jp
Ryohei Egusa, Kyoto Women's University, egusar@kyoto-wu.ac.jp
Satoshi Takahashi, Nagoya University, takahashi.satoshi.h0@f.mail.nagoya-u.ac.jp

Abstract: This study contributes to computer-supported collaborative learning research by examining human–artificial intelligence (AI) dialogue patterns across teams with different experience levels with generative AI (GAI)-supported inquiry-based learning. Epistemic network analysis of dialogues from 26 teams reveals that teams using GAI support for the second time exhibit projective actions toward their goals through active questioning and concept clarification. In contrast, inexperienced teams rely on GAI scaffolding with passive responses. Notably, goal-oriented questions emerge as critical mediators connecting different question types among first-time user teams.

Background and research question

Inquiry-based learning is a prominent theme in the learning sciences and computer-supported collaborative learning (CSCL). Edelson et al. (1999) identified key learning opportunities, including developing inquiry skills and deepening scientific understanding. Established CSCL environments such as the Knowledge Forum (Scardamalia & Bereiter, 2022) and WISE (Linn et al., 2003) have engaged students in scientific inquiry. Subsequent work has addressed assessment (Gobert et al., 2013) and questioning skills (Hakkarainen & Sintonen, 2002). However, these studies have predominantly focused on inquiry in science education. Keifert and Stevens (2018) broadened this scope, using ethnographic analysis to show that inquiries in daily life consist of three elements: orienting to inquiry, progressing through meaning-making resources, and orienting to closure.

Japanese national education standards (MEXT, 2017a, 2017b, 2018) emphasize collaborative inquiry-based learning from elementary school through high school, structuring the process into four phases: problem identification, information gathering, organization and analysis, and summary and expression. This framework integrates both Keifert and Stevens's (2018) model of question emergence in inquiries in daily life and Edelson et al.'s (1999) model of general inquiry skill development. At the high school level, problem identification relates to student identities and life goals (MEXT, 2018), requiring support to formulate meaningful questions, which Scardamalia and Bereiter (1992) classified as knowledge-based questions rooted in deep personal interests. However, supporting the formulation of such questions remains a significant challenge in practice for high school teachers. As of 2025, 230 selected "super science high schools (SSHs)" had implemented inquiry-based research courses (JST, 2025), requiring teachers to manage approximately 10 teams with varying themes. This instructional challenge demands adaptive support, suggesting the potential of generative artificial intelligence (GAI)-based interactive assistance (Borge et al., 2024; Kasneci et al., 2023).

GAI-supported learning is advancing rapidly in both practice and research, with growing attention being devoted to human–AI collaboration. Ojeda-Ramirez et al. (2025) suggested that student reflection prompts a critical awareness of AI's role in collaborative creative tasks. Ohsaki and Matsukawa (2025) proposed quantitative and visual analyses of AI–student dialogues in inquiry-based learning, demonstrating that commercial GAI often provides excessive support. However, limited attention has been paid to the relationship between learner agency in inquiry and autonomous AI use, despite CSCL research on knowledge creation highlighting shared epistemic agency within teams (Damşa et al., 2010; Scardamalia, 2002).

Accordingly, this study examines how human–AI interaction patterns vary across teams with different experience levels with GAI-supported inquiry to inform CSCL design in the AI era. Specifically, we address the following question: How do human–AI interaction patterns differ between first-time and second-time GAI user teams?

Methods

The GAI tool used in this study was developed through iterative co-design with experienced teachers and guided by five design principles: operational simplicity, automatic logging, explicit mode switching, institutional data storage (Google Workspace), and step-by-step questioning. Structurally, the tool is designed for one computer per team to facilitate group discussion and consists of two modes. The inquiry support mode engages students in adaptive dialogue using the 5W1H framework, drawing on active listening principles (Rogers, 1957) to facilitate

 © ISLS

student-driven inquiry, rather than providing answers directly. The upper-grade research exploration mode enables students to identify upper-grade presentations for their inquiries. Both modes run on either Claude 3.5 Sonnet (Anthropic, 2024) or GPT-4.1 (OpenAI, 2025), and teachers selected the model for their class.

Data were collected during a single lesson in a high school research project course (SSH designation), where 268 students across 87 teams explored self-generated questions. Among the 87 teams, 69 used an AI support tool, with 17 having prior experience. A survey conducted a week later yielded 199 valid responses from 221 students across 85 teams. Each team's responses were summarized using median scores from three five-point Likert items: two measures of GAI support tool effectiveness (research theme specification and motivational impact) and a measure of overall satisfaction. Based on the responses, teams with two or more respondents and a median score of at least four were selected (n = 39) to ensure meaningful engagement with the GAI-assisted tool, reflecting high perceived effectiveness and satisfaction. Among the 26 teams with usable log data, five were second-time users (Group A), and 21 were first-time users (Group B).

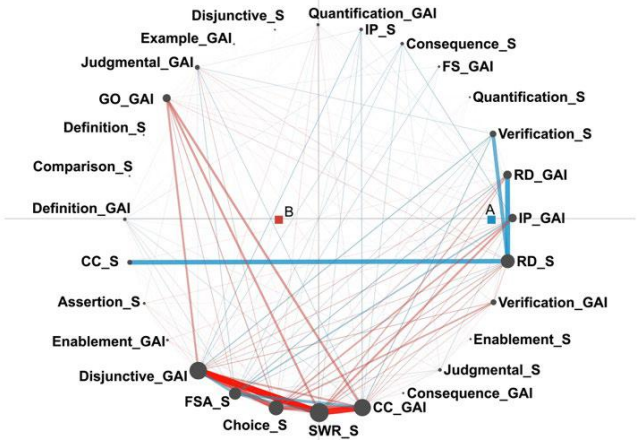
We employed epistemic network analysis (ENA; Shaffer, 2017) to compare interaction patterns between Groups A and B, coding utterances using a 21-code scheme. Eighteen codes were adapted from Graesser and Person (1994). Codes 1 to 10 correspond to basic questions: Verification (yes/no), Disjunctive (either/or options), Concept Completion (CC; who/what), Feature Specification (FS; characteristics), Quantification (quantities), Definition, Comparison, Example, Interpretation, and Judgmental (value-based). Codes 11 to 16 correspond to deeper questions: Antecedent (causes), Consequence (hypothetical outcomes), Goal Orientation (GO; motives), Enablement (means), Instrumental/Procedural (IP; alternative methods), and Expectational (why expected events did not occur). Codes 17 and 18 contained non-question utterances: Assertion (expressing lack of knowledge) and Request/Directive (RD: requesting actions). Codes 19 to 21 were developed to classify student responses: Choice (selecting from given options), Single-Word Response (SWR), and Full Sentence Answer (FSA). Speaker identity was encoded by appending "_GAI" and "_S" suffixes to each code. These coded units were then entered into ENA, configured with Group and Team as the analysis units, a moving stanza window of four, and nodes arranged on a unit circle.

Results and discussion

The ENA results reveal distinct human–AI interaction patterns between first-time and second-time GAI user teams (Figure 1). Stronger connections indicate more frequent co-occurrences of coded lines. The results suggest that collaboration patterns with GAI may vary according to team experience with GAI-supported inquiry and that GAI has the potential to compensate for such differences dynamically. Group A (blue) demonstrated active engagement, as students posed questions and made requests to the GAI (Verification_S, CC_S, RD_S) and provided full-sentence responses (FSA_S), indicating clearer goal recognition and knowledge-based questioning (Scardamalia & Bereiter, 1992). In contrast, Group B (red) exhibited more passive patterns, providing SWRs (SWR_S) or selecting GAI-provided options (Choice_S). Furthermore, the results reveal that GAI's verbalization of student motivation (GO_GAI), corresponding to the "why" question in the 5W1H framework, serves an important bridging function when inquiry goals are unclear, consistent with Goal Orientation as a deep question (Graesser & Person, 1994). Additionally, although not a planned feature, the GAI may have adjusted scaffolding intensity in response to student inputs, suggesting its potential for tailored support in open-ended inquiry.

Figure 1

ENA Analysis Results



 © ISLS

References

Anthropic (2024) Introducing Claude 3.5 Sonnet. https://www.anthropic.com/news/claude-3-5-sonnet

Borge, M., Smith, B.K. & Aldemir, T. (2024). Using generative ai as a simulation to support higher-order thinking. International Journal of Computer-Supported Collaborative Learning.

Edelson, D. C., Gordin, D. N., & Pea, R. D. (1999). Addressing the Challenges of Inquiry-Based Learning Through Technology and Curriculum Design. Journal of the Learning Sciences, 8(3–4), 391–450.

Damşa, C. I., Kirschner, P. A., Andriessen, J. E. B., Erkens, G., & Sins, P. H. M. (2010). Shared epistemic agency: An empirical study of an emergent construct. Journal of the Learning Sciences, 19(2), 143–186.

Gobert, J. D., Sao Pedro, M., Raziuddin, J., & Baker, R. S. (2013). From Log Files to Assessment Metrics: Measuring Students' Science Inquiry Skills Using Educational Data Mining. Journal of the Learning Sciences, 22(4), 521–563. https://doi.org/10.1080/10508406.2013.837391

Graesser, A. C., & Person, N. K. (1994). Question asking during tutoring. American educational research journal, 31(1), 104-137.

Hakkarainen, K., & Sintonen, M. (2002). The Interrogative Model of Inquiry and Computer-Supported Collaborative Learning. Science & Education, 11(1), 25–43. https://doi.org/10.1023/A:1013076706416

JST (Japan Science and Technology Agency) (2025). Super Science High School. https://www.jst.go.jp/cpse/ssh/pamphlet/sshpamphlet.pdf (in Japanese)

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., Stadler, M., Weller, J., Kuhn, J., Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Keifert, D., & Stevens, R. (2018). Inquiry as a Members' Phenomenon: Young Children as Competent Inquirers. Journal of the Learning Sciences, 28(2), 240–278. https://doi.org/10.1080/10508406.2018.1528448

Linn, M. C., Clark, D., & Slotta, J. D. (2003). WISE design for knowledge integration. Science Education, 87(4), 517–538. https://doi.org/10.1002/sce.10086

MEXT (Ministry of Education, Culture, Sports, Science and Technology) (2017a). Commentary on the Course of Study for Elementary School [the period for integrated studies]. https://www.mext.go.jp/component/a_menu/education/micro_detail/__icsFiles/afieldfile/2019/03/18/1387017_013_1.pdf (in Japanese)

MEXT (Ministry of Education, Culture, Sports, Science and Technology) (2017b). Commentary on the Course of Study for Junior High School [the period for integrated studies]. https://www.mext.go.jp/component/a_menu/education/micro_detail/__icsFiles/afieldfile/2019/03/18/1387018_012.pdf (in Japanese)

MEXT (Ministry of Education, Culture, Sports, Science and Technology) (2018) Commentary on the Course of Study for High Schools [Period for Inquiry-Based Cross-Disciplinary Study]. https://www.mext.go.jp/content/1407196_21_1_1_2.pdf (in Japanese)

Ohsaki, A. & Matsukawa, H. (2025). Interaction Analysis Between Generative Ai and High School Students in Scientific Projects. In Oshima, J., Chen, B., Vogel, F., & Järvelä, J. (Eds.), Proceedings of the 18th International Conference on Computer-Supported Collaborative Learning - CSCL 2025 (pp. 425-429). International Society of the Learning Sciences.

Ojeda-Ramirez, S., Han, A., Rich, A., Han, S., & Peppler, K. (2025). Reflective Ai-Partnerships: How Middle Schoolers Balance Creativity and Ai Collaboration. In Oshima, J., Chen, B., Vogel, F., & Järvelä, J. (Eds.), Proceedings of the 18th International Conference on Computer-Supported Collaborative Learning - CSCL 2025 (pp. 404-408). International Society of the Learning Sciences.

OpenAI. (2025). Introducing GPT-4.1 in the API. https://openai.com/index/gpt-4-1/

Rogers, C. R. (1957). The necessary and sufficient conditions of therapeutic personality change. Journal of Consulting Psychology, 21 (2), 95-103.

Scardamalia, M. (2002). Collective cognitive responsibility for the advancement of knowledge. In B. Smith (Ed.), Liberal education in a knowledge society (pp. 67–98). Chicago, IL: Open Court.

Scardamalia, M., & Bereiter, C. (1992). Text-based and knowledge-based questioning by children. Cognition and Instruction, 9, 177–199.

Scardamalia, M., & Bereiter, C. (2022). Knowledge building and knowledge creation. In R. K. Sawyer (Ed.), The Cambridge handbook of the learning sciences (2nd ed., pp. 385-405). Cambridge University Press.

Shaffer, D. W. (2017). Quantitative ethnography. Madison, WI: Cathcart Press.

Acknowledgments

This work was supported in part by JSPS KAKENHI Grant Numbers JP23K11357, JP24K00445, JP25K17107, JP25K06648, and JP25K21929.

International Society of the Learning Sciences

Between Modalities: Observing How Learners Move between Verbal and Digital Spaces in Face-to-Face Collaborative Knowledge Building

Litong Zeng, University of Illinois at Urbana Champaign, litongz2@illinois.edu
Yiqi Xiao, University of California Irvine, yiqix7@uci.edu
Mike Tissenbaum, University of Illinois at Urbana Champaign, miketissenbaum@gmail.com

Abstract: This study explores how learners move between verbal and digital modalities during face-to-face collaborative knowledge building with the Idea Wall, a spatially manipulable digital canvas. Using qualitative multimodal interaction analysis, we examined a ten minute small group session where middle school students jointly discussed pollution. We found that verbal narration often amplifies digital actions, supporting joint attention and shared understanding. Learners frequently moved between verbal and digital communication, however, misalignments occasionally occurred when verbal and digital exchanges failed to synchronize, creating temporary attention gaps. These findings help us understand both the affordances and tensions of multimodal collaboration, and inform our ongoing design-based research on refining the Idea Wall for smoother cross-modality collaboration.

Introduction

In face-to-face collaborative knowledge building (KB) environments, digital tools, particularly interactive surfaces such as digital walls and tabletops (Tissembaum & Slotta, 2019), enrich the physical space by providing shared platforms for organizing and visualizing ideas. This integration enables learners to externalize their thinking, sustain joint attention, and manipulate ideas collaboratively in real time (Zeng et al., 2023), supporting more tangible engagement with abstract concepts. In such settings, intelligence becomes distributed across people, artifacts, and representations, enabling collaborative sensemaking that transcends individual cognition (Pea, 1993; Hollan et al., 2000). But when verbal, digital, and embodied modalities coexist in the same space, learners must coordinate attention and actions across them, creating both opportunities and tensions in collaborative knowledge building depending on how they bridge these engagements (Sung & Nathan, 2025). Research shows that while such tools make collective knowledge visible, they can also fragment attention when learners fail to bridge verbal and digital spaces (Walkington et al., 2019; Schneider et al., 2021; Cheung, 2022). To better understand these dynamics, this paper investigates Idea Wall, a spatially manipulable digital canvas used on tablets, and asks: how do learners move between talk and digital action in face-to-face collaboration, and how do they transfer and align ideas across modalities during collaborative knowledge building?

Method

Participants and Activity

The participants were 12 middle school students from a public STEM magnet school in the U.S. Midwest. Students worked in three groups of three to four. Data sources included system logs, group video, and individual audio and screen recordings. We focus on one group's first-day session as it's representational. The activity unfolded over four lunch periods during a regular school week. On the first day, students investigated pollution through a scaffolded activity introducing the Idea Wall and its collaborative knowledge-building affordances. The interface includes a blank zone for spatially manipulate ideas and add new ideas, collective agreement (Yup Zone), and idea integration and advancement (Combine Zone). Before the activity, students individually wrote pollution examples in the digital notebook. In the "Combine Examples of Pollution" task, they were asked to merge similar notes first, and then connected related examples on the Idea Wall to generate higher-level ideas with the Combine Zone function.

Analysis

This study employs a qualitative, multimodal interaction analysis approach (Norris, 2004). The analysis draws on systematic observation and detailed analytic note-taking. The coding framework used in this study was from the author's earlier work (Zeng et al, 2023), which distinguishes between verbal(*v*) and digital(*d*) modalities across key categories of collaborative knowledge-building actions. Idea Generation *(vIG/dIG)* involves contributing new thoughts orally or via digital notes; Joint Attention *(vJA/dJA)* entails verbally or using digital gestures to draw

 © ISLS

awareness; Making Suggestions *(vMS/dMS)* consists of spoken proposals or moving notes across functional zones; Clarification and Negotiation *(vCN/dCN)* occurs through verbal justification or repositioning notes to signal revision; Rise Above *(vR/dR)* represents high-level synthesis, achieved by articulating integrated understandings or digitally restructuring ideas in the Combine Zone; Narration *(N)* is when students verbalize something existing (only in verbal).

Findings and Discussion

The 10 minutes focal episode involved 83 verbal turns and 54 digital actions. Throughout the activity, students constantly shifted between verbal and digital modalities, using both speech and digital manipulation to negotiate their ideas. The following sections analyze the specific ways these two modalities interacted and how students coordinated their actions to maintain a shared understanding.

Silent Acceptance and Ambiguity of Digital Negotiation

In multimodal knowledge building, digital actions often serve as silent yet definitive communicative turns, where acceptance must be inferred from the movement of artifacts rather than confirmed through speech. A clear instance occurred when P2 initiated a suggestion *(vMS)*, "So microplastic?", while P3 simultaneously moved the note digitally *(dMS)*. This lack of synchronization initially caused a modality clash; as P2 moved the note to the Yup Zone while narrating *(N/dMS)*, P3 lost track of the artifact ("Where did it go?", *vCN/vJA)*. However, following P3's final instruction ("Put the... put in the Combine Zone" *vMS*), P2 offered no verbal reply but immediately moved the note *(dAS)*. Aligning with P3's proposal, this silent move represents a rare moment of clarity where the digital action functioned as an implicit acceptance that completed the negotiation. Yet, such clarity was not the norm. Frequently, notes were moved rapidly and silently, such as from the Yup Zone to the Blank Zone, without verbal framing. Because digital manipulation inherently carries epistemic meaning within the Idea Wall, these unspoken moves create interpretive ambiguity, leaving it unclear whether they represent active agreement, passive acceptance, or simply unnoticed changes. This underscores a core challenge: when actions replace words, understanding becomes inferred rather than confirmed. This reliance on inference could potentially mask underlying conceptual misalignments, suggesting that the benefits of digital fluidity should be weighed against the explicit semiotic clarity needed for sustained collective knowledge building.

Joint Attention to help transition between modalities

Beyond the challenge of understanding what a move means, the digital tool can also split the group's focus. When multiple modalities coexist, learners need to continuously shift their attention and actions between verbal and digital spaces. The digital medium, while smooth and responsive, can absorb the learners' attention. The imbalance becomes more pronounced when group members differ in their preferred modes of communication; some primarily talk, while others interact digitally. In some cases, modality transitions require additional cognitive effort. A question posed verbally may be answered digitally, but the speaker might not notice, leaving a temporary gap in communication. For example, at [14:21], P2 prompted the group, "Let's think, guys… what could I add that's same as oil?"*(vJA/vMS)* While P2 and P3 continued to debate the idea verbally, P1 answered at [15:00] by quietly adding a note for "fossil fuel" without comment *(dIG)*. However, because this move was silent, at [15:04], P3 was still wondering aloud about what to add, completely unaware that the answer was already on the screen. There was no verbal acknowledgement even at the end of the session that P2 and P3 was aware of the new idea P1 generated, only digital agreement. This unsynchronized coordination illustrates how verbal and digital actions can proceed in parallel rather than in full alignment. Digital collaboration, in this sense, carries a silent yet fluid quality, interaction flows smoothly, but shared understanding must be continuously re-established across modalities. Joint Attention (waving the digital notes or narrating actions), most of the time can help transition peers' focus between modalities. For instance, as P2 narrated, "I'm gonna put food waste," he simultaneously added a corresponding note, while P3 responded verbally with "Gas" and "Woo gas, woo type stuff." Speaking and acting digitally worked together, the verbal narration amplified the digital move, making it audible and visible to others. This dual-channel expression not only supported joint attention but also expanded the communicative reach of each action. Knowledge building in this sense emerged through the fusion of modalities, where ideas were co-constructed across what was said, done, and seen.

Overall, these findings suggest that moving between talk and digital action in face-to-face KB is a negotiated process of navigating attentional gaps between group members. Instead of focusing solely on frictionless interaction, perhaps designing for "collaborative seams" could help learners transform silent moves into shared social markers through explicit joint attention. Such an approach might better allow individual digital actions to be anchored and aligned within the group's collective knowledge building process.

References

Cheung, A. (2022). Verbal and on-screen peer interactions of EFL learners during multimodal collaborative writing: A multiple case-study. *Journal of second language writing*, *58*, 100931.

Hollan, J., Hutchins, E., & Kirsh, D. (2000). Distributed cognition: toward a new foundation for human-computer interaction research. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *7*(2), 174-196.

Norris, S. (2004). *Analyzing multimodal interaction: A methodological framework*. Routledge.

Pea, R. D. (1993). Practices of distributed intelligence and designs for education. *Distributed cognitions: Psychological and educational considerations*, *11*, 47-87.

Schneider, B., Worsley, M., & Martinez-Maldonado, R. (2021). Gesture and gaze: Multimodal data in dyadic interactions. In *International handbook of computer-supported collaborative learning* (pp. 625-641). Cham: Springer International Publishing.

Sung, H., & Nathan, M. J. (2025). Unraveling temporally entangled multimodal interactions: investigating verbal and nonverbal contributions to collaborative construction of embodied math knowledge. *International Journal of Educational Technology in Higher Education*, *22*(1), 8.

Tissenbaum, M., & Slotta, J. D. (2019). Developing a smart classroom infrastructure to support real-time student collaboration and inquiry: a 4-year design study. *Instructional Science*, *47*(4), 423-462.

Walkington, C., Chelule, G., Woods, D., & Nathan, M. J. (2019). Collaborative gesture as a case of extended mathematical cognition. *The Journal of Mathematical Behavior*, *55*, 100683.

Zeng, L., Hadinezhad, S., & Tissenbaum, M. (2023, June). Spatially Manipulable Interactive Note Taking Tool to Facilitate Collaborative Learning in Science Education. In *Proceedings of the 22nd Annual ACM Interaction Design and Children Conference* (pp. 614-618).

Acknowledgements

This paper used ChatGPT-5 to assist with academic writing and language refinement. All ideas, interpretations, and arguments presented are original and authored by the researcher.

International Society of the Learning Sciences

One Size Fits All? Learner-Adaptive Peer Feedback and Perceived Feedback Adequacy in Pre-Service Teacher Education

Alina Sheppard-Bujtor, Martin Greisel, Julia Hornstein, Ingo Kollar
alina.sheppard-bujtor@uni-a.de, martin.greisel@uni-a.de, julia.hornstein@uni-a.de, ingo.kollar@uni-a.de
University of Augsburg

Abstract: Peer feedback can effectively facilitate pre-service teachers' evidence-informed reasoning in CSCL environments. Based on the RESOLV-model, we examine if peer feedback providers plan and formulate their peer feedback messages differently depending on the perceived skill level of the peer feedback recipient. This will be analyzed by assessing the feedback feature constellations of feedback provided by *N*=297 pre-service teachers during a psychology course. We expect that the more feedback providers adapt their feedback, the more positively this will affect recipients' acceptance.

Aims and theoretical background

Online peer feedback is widely used in the computer-supported collaborative learning (CSCL) context to foster learning outcomes (Tan & Chen, 2022). It has great potential as a pedagogical approach (Double et al., 2020), as it structures peer interaction in a way that ensures transactivity (Vogel et al., 2016): Optimally, students thoroughly process their peers' contributions, assess them, react to them, and build upon them to extend their peers', but also their own understanding of subject-matter information (Yu & Schunn, 2023).

Yet, feedback recipients do not necessarily benefit from peer feedback (Li et al., 2020). While it is crucial for recipients to understand their peer feedback in order to learn from it (Nelson & Schunn, 2009), they can get overwhelmed by, for example, very comprehensive and complex feedback which addresses multiple interrelated points (Shulgina et al., 2024). This could be especially challenging for recipients with limited prior knowledge of the topic. To mitigate this, feedback providers need to adapt the scope and depth of their feedback message to the recipient's skill level. If they do so, this should positively affect the recipient's assessment of the adequacy of the feedback, leading to more likely implementation of the feedback in the subsequent revision, and ultimately to better learning outcomes (Strijbos et al., 2021). Accordingly, more adaptive peer feedback messages are expected to improve learning outcomes and are therefore relevant within the CSCL context.

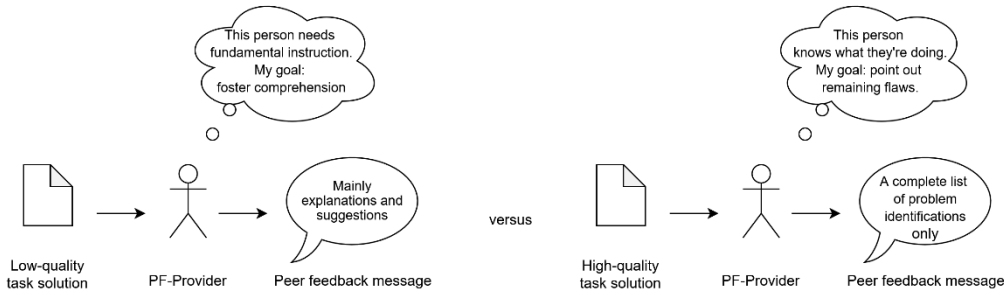
Surprisingly, prior CSCL research seems to assume that certain feedback features (such as the presence of explanations and solutions) work in a one-size-fits-all fashion (to be generally beneficial for subsequent learning, i.e., independently from the recipient's skill level; e.g., Wu & Schunn, 2020). Yet, in line with research on the expertise-reversal effect (Kalyuga et al., 2003), we assume that the relative importance of feedback features—such as explanations and solutions, which are deemed essential to the recipient's learning outcome (Wu & Schunn, 2020)—depends on the recipient's skill level, as evidenced by the quality of their initial task solution performance. Specifically, detailed explanations/solutions would be more important for recipients who have not yet mastered the task as successfully as more competent recipients. In turn, these more competent learners likely require only identification of problems in their solutions. Ideally, the feedback provider should be able to recognize these differences in the recipient's skill levels and tailor their feedback message accordingly. Based on the RESOLV-model, the feedback provider would evaluate the author's expertise (Rouet et al., 2017). The feedback provider judges the level of competence exhibited via the feedback recipient's task performance and based on this, recognizes their skill level. Next, the feedback provider would formulate a goal and a plan on how to adjust their message to match the recipient's skill level. This aligns with Zimmermann's (2002) forethought phase, in which the feedback provider sets the goal to adapt their feedback message to the level of competence they assume the recipient to have. Thereby, feedback providers balance their message between completeness and ease of comprehension. We assume that these kinds of cognitive processes determine the proportion of feedback features (in this case: solutions and explanations) in the feedback message (see Fig. 1).

Hypotheses

We assume that university students are able to adapt their feedback to their peers' skills levels, at least to a certain extent. We therefore hypothesize that the worse the recipient's initial performance is, the more explanations and solutions the provider's feedback contains (H1), as illustrated in Figure 1. Furthermore, we expect that the more the provider adapts their feedback to the initial performance, the more adequate the recipient perceives it (H2).

 © ISLS

Figure 1
How an Adaptive Peer Reviewer Would React to a Low- vs. High-Quality Task Solution



Method

The sample consists of N = 297 pre-service teachers of differing subjects, most of which were women (74.05%) with an average age of 22.26 years (SD = 2.76). Most students were in their fifth semester of studies (M = 5.21, SD = 1.66). The participants engaged in an asynchronous online peer feedback activity over the duration of four weeks, embedded in parallel psychology courses. The course goal was to develop evidence-informed reasoning skills. Therefore, students analyzed problematic classroom situations utilizing educational and psychological theories following an evidence-informed reasoning scheme (problem identification, description, explanation, goal derivation, and action plan; Greisel et al., 2022). In the university's learning management system, the students submitted their work and collaborated via the computer-supported peer feedback tool 'getFeedback!' (Fey et al., 2023), which structures the peer feedback process from start to finish. This encompasses the provision of instructions, scaffolding, assignment of peer feedback pairs, and delivering feedback messages. In Week 1, students crafted an initial task performance by analyzing problematic classroom situations. In Week 2, they provided feedback on the initial task performance of two randomly assigned peers (scaffolded through content-related guiding questions and a corresponding worked example). In Week 3, students revised their feedback. In Week 4, students received feedback from two peers and revised their initial task performance.

Two independent coders assessed the quality of the initial task performance regarding evidence-informed reasoning skill level according to Greisel et al. (2022) with an interrater-reliability of Gwet's $AC2$ = .84. To assess the presence of feedback features, the provided messages were segmented into topical segments (Gwet's $AC1$ = .82) according to the expected evidence-informed reasoning scheme. The features will then be coded depending on their presence, resulting in dummy coded variables, based on a scheme by Wu and Schunn (2020). Coded features encompass: Praise, identification, explanation, hedges, suggestion, solution, accuracy, and importance. The coding process is currently under development, and coder training will aim for an interrater-reliability > .80. Feedback adequacy was assessed through the Feedback Perception Questionnaire by Strijbos et al. (2021), for each of the two feedback messages separately, encompassing three dimensions (fairness, utility, acceptance) measured on a Likert-scale ranging from 1 = *completely disagree* to 9 = *completely agree* with McDonald's ω_h = .82/.85.

Multilevel regression analyses will be applied to test the effects of evidence-informed reasoning skill level of the initial task performance on the relative proportion of solutions and explanations, that is, the presence of explanations and solutions divided by the number of all segments. With exploratory cluster analyses, we will further investigate whether combinations of feedback features are influenced by the task performance level. Feedback adequacy will be analyzed through analyses of variances, accounting for the adaptivity of feedback messages.

Potential results and implications

As the coding process is still ongoing, we cannot yet present results. Should providers formulate their feedback adaptively as predicted and thereby foster a more adequate perception of the feedback message, the one-size-fits-all assumption of peer feedback quality needs to be rethought. Ideally, scaffolding for feedback providers could be made adaptive to initial performance levels. For example, peer feedback systems, like 'getFeedback!', could tailor whether they prompt the feedback provider to explain in accordance with the initial performance level. This could improve the learning experience and outcomes for both the feedback providers and the recipients (Kalyuga et al., 2003; Nelson & Schunn, 2009). In case of non-significant findings, a different (e.g., qualitative) approach to adaptivity assessment may be more fruitful.

 © ISLS

References

Double, K. S., McGrane, J. A., & Hopfenbeck, T. N. (2020). The impact of peer assessment on academic performance: A meta-analysis of control group studies. *Educational Psychology Review*, *32*(2), 481–509. https://doi.org/10.1007/s10648-019-09510-3

Fey, C.-C., Wekerle, C., Beckmann, I., Schröder, A., & Kollar, I. (2023). Agile Methoden in Entwicklungsprojekten zur Innovation digitaler Hochschullehre. [Agile methods in developmentprojects for the innovation of digital higher education]. *Zeitschrift Für Hochschulentwicklung*, *18*(3), 275–294. https://doi.org/10.21240/zfhe/18-03/14

Greisel, M., Wekerle, C., Wilkes, T., Stark, R., & Kollar, I. (2022). Pre-service teachers' evidence-informed reasoning: Do attitudes, subjective Norms, and self-efficacy facilitate the use of scientific theories to analyze teaching problems? *Psychology Learning & Teaching*, *22*(1), 20-38. https://doi.org/10.1177/14757257221113942 (Original work published 2023)

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, *38*(1), 23–31. https://doi.org/10.1207/S15326985EP3801_4

Li, H., Xiong, Y., Hunter, C. V., Guo, X., & Tywoniw, R. (2020). Does peer assessment promote student learning? A meta-analysis. *Assessment & Evaluation in Higher Education*, *45*(2), 193–211. https://doi.org/10.1080/02602938.2019.1620679

Nelson, M. M., & Schunn, C. D. (2009). The nature of feedback: How different types of peer feedback affect writing performance. *Instructional Science*, 37(4), 375–401. https://doi.org/10.1007/s11251-008-9053-x

Rouet, J.-F., Britt, M. A., & Durik, A. M. (2017). RESOLV: Readers' representation of reading contexts and tasks. *Educational Psychologist*, *52*(3), 200–215. https://doi.org/10.1080/00461520.2017.1329015

Shulgina, G., Fanguy, M., Zhang, H., Courtney, M., Baldwin, M., & Costley, J. (2024). The moderating effects of total comments on the relationship between comment implementation and online peer-supported writing performance. *Computers & Education*, *219*, 105104. https://doi.org/10.1016/j.compedu.2024.105104

Strijbos, J.-W., Pat-El, R., & Narciss, S. (2021). Structural validity and invariance of the Feedback Perceptions Questionnaire. *Studies in Educational Evaluation*, *68*, 100980. https://doi.org/10.1016/j.stueduc.2021.100980

Tan, J. SH., & Chen, W. (2022). Peer feedback to support collaborative knowledge improvement: What kind of feedback feed-forward? *Computers & Education*, *187*, 104467. https://doi.org/10.1016/j.compedu.2022.104467

Vogel, F., Kollar, I., Ufer, S., Reichersdorfer, E., Reiss, K., & Fischer, F. (2016). Developing argumentation skills in mathematics through computer-supported collaborative learning: The role of transactivity. *Instructional Science*, *44*(5), 477–500. https://doi.org/10.1007/s11251-016-9380-2

Wu, Y., & Schunn, C. D. (2020). From feedback to revisions: Effects of feedback features and perceptions. *Contemporary Educational Psychology*, *60*, 101826. https://doi.org/10.1016/j.cedpsych.2019.101826

Yu, Q., & Schunn, C. D. (2023). Understanding the what and when of peer feedback benefits for performance and transfer. *Computers in Human Behavior*, 147, 107857. https://doi.org/10.1016/j.chb.2023.107857

Zimmerman, B. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice*. *41*. 64-70. https://doi.org/10.1207/s15430421tip4102_2

International Society of the Learning Sciences

If One Hack Week is Pedagogically Effective, Why Not Ten?

Jacob Puthipiroj, Northwestern University, puthipiroj@u.northwestern.edu
Thepchai Supnithi, NECTEC, thepchai.supnithi@nectec.or.th
Nongnuch Ketui, Rajamangala University of Technology Lanna, nongnuchketui@rmutl.ac.th

Abstract: While hackathons are established pedagogical tools, questions persist about measuring their impact when integrated longitudinally into curricula. We investigate a data science curriculum incorporating eight Kaggle-based hackathons over ten weeks with 137 participants. Using immersive ethnographic methods, we constructed narrative vignettes illuminating how longitudinal hackathon participation fosters identity development, distributed cognition, and community formation. Drawing on Communities of Practice and distributed cognition as both design principles and analytic lenses, we argue that hackathons, when embedded as recurring curricular structures, function as sites for constructing consequential learning identities.

Introduction

Hackathons are valued for fostering experiential learning, collaboration, and professional skills (Towhidi, 2022; Flus, 2022; Oyetade, 2024), yet fundamental challenges persist: most research remains empirical rather than theoretical (Schulten, 2024), assessment relies on self-reported surveys, and studies predominantly examine hackathons as isolated events rather than longitudinal curricular structures (Chau & Gerber, 2023). We address these gaps by examining a data science curriculum in which hackathons serve as the primary pedagogical method over ten weeks with 137 participants in eight Kaggle-based competitions. Unlike Puthipiroj (2025), who quantitatively assesses outcomes through rankings and submission frequency, we adopt immersive ethnography to capture the experiential and identity-forming dimensions of sustained hackathon participation.

Our research questions are: (1) How does sustained participation in a longitudinal hackathon curriculum shape participants' learning identities and community membership? (2) How are cognitive and collaborative processes distributed across people, tools, and social structures during extended hackathon engagement?

Program design

During early 2025, an AI/ML program selected 137 participants from an open virtual first phase into an intensive second phase. Participants were divided into six teams of roughly 25, composed to balance experience levels—mixing high school students (ages 15–18), undergraduates, recent graduates, and working professionals (up to age 50)—with varying proportions of programming backgrounds. The program featured eight five-day Kaggle hackathons: two weeks online via Discord, four weeks onsite at a residential site in Thailand where participants lived alongside mentors, and two final online weeks. Each weekly cycle introduced a Kaggle challenge on Monday, provided lectures Tuesday–Wednesday, and devoted Thursday–Friday to development. Kaggle's real-time leaderboard provided immediate feedback, and weekly pitching rounds evaluated originality and communication beyond predictive accuracy.

Theoretical framework

Two frameworks guided both design and analysis. Communities of Practice (Lave & Wenger, 1991) frames learning as increasing participation in a community; we designed the curriculum to scaffold legitimate peripheral participation and analytically traced how participants moved from peripheral to central roles, attending to moments when community membership became salient. Distributed cognition (Salomon, 1993) provided a lens for understanding how cognitive work was shared across team members, tools (Kaggle, LLMs, Discord), and social settings; in design this motivated the team-based, multi-platform structure, and analytically we traced how intellectual labor was distributed during critical problem-solving episodes.

Methods

The first author conducted immersive ethnographic fieldwork throughout the program—residing in the same hotel during the onsite phase, visiting every team room daily, and documenting observations in fieldnotes covering team dynamics, task allocation, and interpersonal interactions. Digital ethnography extended this presence through Discord and Gather Town during online phases. Twenty-four semi-structured interviews (30–60 minutes) captured perspectives across the full demographic range (ages 15–50). Fieldnotes, transcripts, and communication

 © ISLS

logs were analyzed through iterative thematic coding oriented by our two frameworks, and recurring patterns were distilled into narrative vignettes built from triangulated data sources to illuminate distinct theoretical dimensions.

Findings

Vignette 1: Distributed cognition under pressure

On Week 3, Jean, a 20-year-old university graduate who had participated in dozens of hackathons and sold medical algorithms to hospitals, took on unofficial leadership of his team. The task, developing forecasting models for retail SKUs, seemed straightforward, but inconsistencies accumulated: missing values without logical patterns, duplicate entries across time periods. During a meeting with the corporate sponsor, Jean asked what his teammates had all been thinking: "Why is your data so poorly kept?" The question violated unspoken protocols of deference. Organizers pulled him aside with concerns, but his teammates defended him. Their investigation revealed that the evaluation metric itself appeared rule-based—a crucial insight. With only four Kaggle submissions allowed per day, each became a costly experiment probing the hidden logic of the ground truth.

Midway through the week, a COVID outbreak forced 11 team members into hotel-room isolation—the most of any team. The team's cognitive network adapted: isolated members remained tethered through Discord, analyzing data and sharing observations. Crucially, an isolated member noticed a pattern in submission scores that the room-based subgroup had missed—reshaping their final strategy. Distributed cognition operated across physical and digital spaces, with the platform mediating contributions no co-located subgroup could have produced alone. Jean's trajectory also illustrates legitimate peripheral participation in reverse: his expertise initially positioned him centrally, but the team's validation of his confrontational approach—and the organizers' pushback—socialized him into the community's norms around challenging authority constructively.

Vignette 2: Identity formation through sustained participation

By the time Week 6 arrived, Pine's stress had reached a maximum. At eighteen, fresh from high school graduation, she was acutely aware of what everyone already knew: her team was fundamentally different. While other groups boasted participants with extensive hackathon experience, polished portfolios, and even a PhD researcher and an adjunct professor, Pine's group had finished last twice and second-to-last three times. Week 6 carried the largest prize pool yet, backed by a major Thai bank actively recruiting for internships.

A team gathering that was supposed to help nearly broke them. Staff and researchers were asked to leave the room, and everyone anonymously wrote down their reasons for attending. Two distinct groups emerged: one vocal faction wanted to win and double their efforts; the other, quieter but equally present, had come primarily for learning—the prize was secondary to their curiosity. This moment—what Lave and Wenger might recognize as the community negotiating its shared enterprise—proved catalytic rather than divisive.

Against all expectations, Pine's team won. What made the victory particularly meaningful was the timing: Week 6 was the final onsite week, and organizers had planned a grand celebration where researchers, lecturers, staff, and all 137 participants gathered to mark the end of the intensive month. As the evening unfolded, each team voted for their most valuable member. Pine heard her name—recognized not for technical prowess but for holding the team together across weeks of losses. In Nasir's (2020) terms, her trajectory exemplifies consequential participation: she moved from peripheral uncertainty to recognized community membership, constructing a learning identity through sustained engagement rather than isolated achievement.

Discussion and conclusion

These vignettes reveal what surveys miss: the formation of what Nasir (2020) describes as consequential participations in learning pathways. Both Jean and Pine underwent identity transformations only legible across weeks of sustained engagement. Pine moved from peripheral uncertainty to recognized community membership; Jean's expertise was reshaped by the community's norms around constructive challenge. These are trajectories that require longitudinal observation to document, not outcomes that post-event satisfaction surveys can detect. Jean's story further illustrates how distributed cognition operates as an emergent property of longitudinal engagement—his team's cognitive network reorganized across physical isolation and digital platforms over days, not hours, with Discord functioning as genuine cognitive infrastructure because the team had built shared practices over prior weeks. Critically, these processes required time—the ten-week structure provided what Nasir calls "the fodder" for constructing productive identities. The field should move beyond evaluating hackathons as isolated interventions and investigate how recurring, longitudinally embedded structures create conditions for consequential learning. Future work should examine how team composition moderates identity formation, and how onsite-to-online transitions reshape distributed cognitive networks.

 © ISLS

References

Chau, C. W., & Gerber, E. M. (2023, April). On hackathons: A multidisciplinary literature review. In *Proceedings of the 2023 CHI conference on human factors in computing systems* (pp. 1-21).

Flus, M., & Hurst, A. (2022). Hackathons as a novel pedagogy in engineering design education. *International Journal of Engineering Education*, 38(1), 36-44.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge university press.

Oyetade, K., Harmse, A., & Zuva, T. (2023). Evaluating student satisfaction with online hackathon for IT projects. *Eurasia Journal of Mathematics, Science and Technology Education*, 19(10), em2338.

Oyetade, K., Zuva, T., & Harmse, A. (2024). Evaluation of the impact of hackathons in education. *Cogent Education*, 11(1), 2392420.

Nasir, N. I. S., de Royston, M. M., Barron, B., Bell, P., Pea, R., Stevens, R., & Goldman, S. (2020). Learning pathways: How learning is culturally organized. In *Handbook of the cultural foundations of learning* (pp. 195-211). Routledge.

Puthipiroj, J., Supnithi, T., Ketui, N., & Thongthungwong, N. (2025, July). Structured Hackathons: A Pedagogical Model For AI Education. In *International Conference on Artificial Intelligence in Education* (pp. 261-273). Cham: Springer Nature Switzerland.

Salomon, G. (Ed.). (1997). Distributed cognitions: Psychological and educational considerations. Cambridge University Press.

Schulten, C., & Chounta, I. A. (2024). How do we learn in and from Hackathons? A systematic literature review. *Education and Information Technologies*, 29(15), 20103-20134.

Towhidi, G., & Pridmore, J. (2022). Hackathons for experiential learning in IS higher education. *Issues in Information Systems*, 23(1).

Acknowledgments

This research has received funding support from the National Science, Research and Innovation Fund (NSRF) via the Program Management Unit for Human Resources & Institutional Development, Research and Innovation [grant number B13F670080].

International Society of the Learning Sciences

Integrating Individualized Education Programs into AI-based Tutoring Systems: An Interview Study with Teachers

C. Naomie Williams, Carnegie Mellon University, Pittsburgh PA 15213, USA, naomiew@cmu.edu
Margaret Ellen Seehorn, University of Michigan, Ann Arbor MI 48104, USA, seehorn@umich.edu
Haiyi Zhu, Carnegie Mellon University, Pittsburgh PA 15213, USA, haiyiz@cmu.edu
Vincent Aleven, Carnegie Mellon University, Pittsburgh PA 15213, USA, aleven@cs.cmu.edu

Abstract: US-based special education teachers use Individualized Education Programs (IEPs) to tailor academic resources for students with disabilities. However, prior interviews reveal significant challenges integrating key IEP components – individualized learning goals, accommodations, and progress monitoring – into existing AI-based tutoring and assessment technologies. This paper discusses AI-based tutoring systems' misaligned analytics with teachers' needs, the level of decision-making AI should have in special education classrooms, and a pathway for adapting AI-based tutoring systems to special education's needs.

Introduction

This works investigates how AI-based tutoring systems influence US-based teachers' practices when personalizing curriculum for students with disabilities. AI systems have promising roles in the future of personalized education for students (Kooli & Chakraoui, 2025; Mohd Amin et al., 2025). However, sole reliance on AI is rare, with teachers actively maintaining separate, preexisting strategies for personalization within their classrooms (Williams et al., 2025). While some prior work explored interventions for specific learning activities and disabilities (Chang, 2022; Garnier et al., 2023; Khamparia et al., 2020; Zingoni et al., 2021), there is a lack of research bridging the functionality of AI-based tutoring systems with the personalization methods teachers actively use to supplement and override AI decisions around students with disabilities (Williams et al., 2025).

We present findings from a series of interviews with special education teachers who use AI-based tutoring systems, aimed at understanding how these tools are used within their pre-existing practices. We explore the following research question: *How might we integrate personalized goals and accommodations into the design and use of AI-based tutoring software in special education classrooms?* We found that current AI-based tutoring systems' priorities do not match the personalized goals within an IEP and fail to offer useful learning analytics. Further, the desired level of AI's decision-making in determining student progress varied among special education teachers. This paper presents preliminary findings from a qualitative investigation of the gap between AI-based tutoring and special education personalization needs.

Related work

Special Education is an inclusive learning environment for students with disabilities. Additional resources in the form of assistive technology, modified curriculum, and personalized learning goals are made possible using Individualized Education Programs (IEPs) (Christle & Yell, 2010). In special education classrooms, students often have multiple disabilities and co-learn with others with a variety of support needs. This environment differs from prior scholarship, which has primarily studied personalized support for specific disabilities in isolation (Allman et al., 2024; Cinquin et al., 2019; Panjwani-Charania & Zhai, 2023; Yang et al., 2024). In this unique environment, special education teachers supplement and override AI decision-making using personalization methods established in individual students' IEPs (Williams et al., 2025). This work contributes to the scholarship bridging theoretical studies to real-world special education classrooms.

Methods

We interviewed 12 special education teachers that both used AI-based tutoring systems in their classrooms and had experience with the IEP process. Five teachers were prior study participants, one teacher was recruited via an education-specific sub-reddit, and the remaining 6 were recruited using Prolific. Teachers who were not previous participants verified their teacher status by providing an official school email address. We assured all teachers that participation was voluntary, responses would be anonymized, and consent could be revoked at any time. To ensure participants remain anonymous, participant data is reported on aggregate with participant IDs.

Semi-structured narrative interviews

 © ISLS

Our interview protocol inquires about the mismatch between special education's methods for personalized instruction and AI-based learning technologies' capabilities. The interview was 1 hour long, containing two parts. The first used directed storytelling to understand how AI-based tutoring and assessment technologies are currently used within the IEP process. *What role do these tools have from data assessment, establishing a student's baseline, and creating measurable, personalized learning goals?* The second half used design scenarios to: 1) Identify teachers' needs, privacy concerns, and desire for control of the availability of their students' IEP information to an AI-based tutoring system used in the classroom. (2) Identify teachers' information needs so they can keep track of personalization that occurs.

Analysis

The interviews were completed and analyzed in two sets of 6. The first six interviews were coded by both coders to establish reliability, the second by only one coder. Each transcript had two passes: one for the big picture ideas and the second for specificity. Finally, we gathered all the codes for a thematic analysis. Using those codes, we completed an affinity diagram, iterating over clusters to identify themes and insights.

Results

Misalignment of software-provided data with teacher's needs

The mismatch of an AI tutoring system's assessment goals with the learning criteria used in Special Education results in poor assessment data utility. *"We don't teach phonics in middle school.* [...], so *we could not write an 8th grader a phonics goal, [...] We are working on comprehension. We are working on writing skills[...]. when you don't pass the phonics [...], those early skills, I can't see how your near or grade level skills are…*", noted P8. However, they expand to attribute poor utility to the lack of data these AI tools provide overall, noting that the software provides primarily a numerical summary, rather than anything useful for analysis of specific skills. "*It might say they got 60% right[...] and you can see they missed number two. But then, I have to on a side sheet of paper, [...] analyze it… they understand pennies, but they don't understand dimes and nickels […], would be nice if it broke everything down, you know."*, expressed P11. Classroom AI provides insufficient detail to support personalized learning goals, leaving teachers with additional labor.

The responsibility of AI varies based on teachers' desired use of the tool

Participant attitudes towards the role of AI in special education generally fall into the following three categories: AI as a tool, AI as a co-teacher, and AI as a collaborator. For AI as a tool, P8 uses AI as a *"data collection instrument to measure [progress towards IEP] goals quarterly"*. AI for this teacher are tools, not decision makers. P12 does not want the AI to make decisions at all, whereas P5 restricts decision making around defining progress to human educators only. In contrast, AI as a co-teacher values AI-provided decisions. The multi-level nature of P10's classroom adds value to AI decisions, where tracking each student individually becomes difficult. To match the independent, self-paced environment, AI decisions need to be trustworthy independent of the teacher. AI having additional knowledge from the IEP brings hope for holistic decisions that apply to multiple subjects; however, those decisions should be transparent. *"It is very important for the teacher to know what decisions got made, because [AI] could contribute to teaching strategies in other areas."* AI as a collaborator combines AI as a tool and AI as a co-teacher. P4 wants AI to *"offer some sort of suggestions on the way forward, but not the actualization itself... final decisions on the actual measures to be taken to define some sort of actual progress should be left for us as educators"*. In a collaborative role, AI should assist, not lead.

Discussion

Our work affirms special education teachers' desire for comprehensive learning analytics and belief thar current AI-based tutoring systems fall short (Nguyen et al., 2018; Williams et al., 2025). We expand upon the reasoning for AI's pitfalls: discussing examples of how assessment data isn't sufficiently granular to be used meaningfully by teachers. Our study contributes new insight into the responsibilities of AI-based tutoring systems in their classrooms, with teachers divided on the level of power AI should have in measuring student's progress. We imagine a future system where special education teachers can provide information from an IEP, such as accessible formats and mastery criteria for learning goals. The software would adjust the default user experience and re-standardize student progress data according to the IEP for easy interpretation by teachers: reducing the burden special education teachers face creating and organizing student personalization data while creating avenues for AI-based tutoring systems to provide more relevant progress tracking, learning analytics, and recommendations.

 © ISLS

References

Allman, B., Kimmons, R., Wang, W., Bao, H., Rosenberg, J. M., & Koehler, M. J. (2024). Trends and Topics in Educational Technology, 2024 Edition. *TechTrends*, *68*(3), 402–410. https://doi.org/10.1007/s11528-024-00950-5

Chang, C.-H. S. (2022). Design and evaluation of a multi-sensory scaffolding gamification science course with mobile technology for learners with total blindness. *Computers in Human Behavior*.

Christle, C. A., & Yell, M. L. (2010). Individualized Education Programs: Legal Requirements and Research Findings. *Exceptionality*, *18*(3), 109–123. https://doi.org/10.1080/09362835.2010.491740

Cinquin, P.-A., Guitton, P., & Sauzéon, H. (2019). Online e-learning and cognitive disabilities: A systematic review. *Computers & Education*, *130*, 152–167. https://doi.org/10.1016/j.compedu.2018.12.004

Garnier, P., Martel, K., Dachez, J., Audry, P., Bourgoin, P., & Stawinski, F. (2023). Educators' perspectives on working with a humanoid robot in a French preschool class for autistic children. *Journal of Research in Special Educational Needs*, *23*(3), 163–174. https://doi.org/10.1111/1471-3802.12588

Khamparia, A., Pandey, B., & Mishra, B. P. (2020). Effects of microworld game-based approach on neuromuscular disabled students learning performance in elementary basic science courses. *Education and Information Technologies*, *25*(5), 3881–3896. https://doi.org/10.1007/s10639-020-10142-2

Kooli, C., & Chakraoui, R. (2025). AI-driven assistive technologies in inclusive education: Benefits, challenges, and policy recommendations. *Sustainable Futures*, *10*, 101042. https://doi.org/10.1016/j.sftr.2025.101042

Mohd Amin, M. R., Ismail, I., & Sivakumaran, V. M. (2025). Revolutionizing Education with Artificial Intelligence (AI)? Challenges, and Implications for Open and Distance Learning (ODL). *Social Sciences & Humanities Open*, *11*, 101308. https://doi.org/10.1016/j.ssaho.2025.101308

Nguyen, A., Gardner, L. A., & Sheridan, D. (2018). A framework for applying learning analytics in serious games for people with intellectual disabilities. *British Journal of Educational Technology*, *49*(4), 673–689. https://doi.org/10.1111/bjet.12625

Panjwani-Charania, S., & Zhai, X. (2023). *AI for Students with Learning Disabilities: A Systematic Review* (SSRN Scholarly Paper No. 4617715). Social Science Research Network. https://papers.ssrn.com/abstract=4617715

Williams, C. N., Seehorn, M. E., Zhu, H., & Aleven, V. (2025). Challenges and Design Opportunities for AI-Based Tutoring and Assessment Software in Special Education: An Interview Study with Teachers. In K. Tammets, S. Sosnovsky, R. Ferreira Mello, G. Pishtari, & T. Nazaretsky (Eds.), *Two Decades of TEL. From Lessons Learnt to Challenges Ahead* (pp. 519–533). Springer Nature Switzerland. https://doi.org/10.1007/978-3-032-03870-8_35

Yang, Y., Chen, L., He, W., Sun, D., & Salas-Pilco, S. Z. (2024). Artificial Intelligence for Enhancing Special Education for K-12: A Decade of Trends, Themes, and Global Insights (2013–2023). *International Journal of Artificial Intelligence in Education*. https://doi.org/10.1007/s40593-024-00422-0

Zingoni, A., Taborri, J., Panetti, V., Bonechi, S., Aparicio-Martínez, P., Pinzi, S., & Calabrò, G. (2021). Investigating Issues and Needs of Dyslexic Students at University: Proof of Concept of an Artificial Intelligence and Virtual Reality-Based Supporting Platform and Preliminary Results. *Applied Sciences*, *11*(10), Article 10. https://doi.org/10.3390/app11104624

 © ISLS

International Society of the Learning Sciences

Computational Modeling of the Process of Scientific Inquiries

Nanae Watanabe, Kyushu University, watanabe.nanae.315@s.kyushu-u.ac.jp
Masahiro Tada, Kindai University, tada@info.kindai.ac.jp
Shin'ichi Konomi, Kyushu University, konomi.shinichi.848@m.kyushu-u.ac.jp
Keisuke Koba, Kyoto University, koba.keisuke.8v@kyoto-u.ac.jp
Masaya Okada, Kyushu University, m.okada@acm.org

Abstract: Scientific inquiry is a process to understand the complex real world. To automatically assess and foster learners' scientific inquiries by a computer-supported collaborative learning (CSCL) system, thinking processes required for scientific inquiries should be modeled in a machine-readable and traceable form. By using a computational modeling approach, the present study extracted patterns of such thinking processes from the data repository of our educational practices and investigated how the processes can be computationally modeled.

Introduction

Scientific inquiry is a process to understand the complex real world. In this process, understanding of the world (e.g., internal representations of the world) should be formatively constructed (National Research Council, 2012). Based on the theory of grounded cognition (Barsalou, 2008), which explains how a person understands the world through real-world experiences, the present study considers that the formative construction process of internal representations involves abstracting information from the world and grounding the representations on real-world experiences. However, even with the latest computer-supported collaborative learning (CSCL) systems, automatically tracing and assessing such process is considered a challenging issue. One of the difficulties is considered to be constructing machine-readable and traceable learner models. By using a computational modeling approach that represents a person's internal information processing as "computation" (Huys et al., 2021), the present study investigates scientific thinking processes that learners should cultivate for formatively constructing the internal representations of the world, and discusses how the processes can be computationally represented.

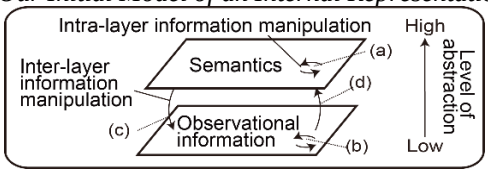
Progressive refinement of a computational model

Since modeling unknown computation processes is generally difficult, this study (i) initially assumed a *qualitative* model of the thinking processes for formatively constructing internal representations, and (ii) progressively refined this model to be expressed in a *computable* form, by using the data repository of our educational practices (e.g., the data of real-world learning in the natural environment, the data of lectures in classroom learning).

In the step (i), our previous study (Watanabe et al., 2023) assumed that a learner performs hierarchical information conversions to update his/her knowledge space which is a hierarchical structure of several information layers with different abstraction levels. Based on this assumption, the previous study made a model that the internal representation is a hierarchical structure where the top layer is semantics (e.g., systematic functions, metrics, model representations) and the bottom layer is observational information, and is constructed through intra-layer information manipulations ((a) and (b) in Fig. 1) and inter-layer information manipulations ((c) and (d) in Fig. 1). However, the model of Fig. 1 did not give computable representations (e.g., a graph representation for numerical calculation) of scientific thinking processes associated with the information manipulations (a)–(d).

Figure 1

Our Initial Model of an Internal Representation of the World



In the step (ii), an important technique of the present paper for investigating how learners should inquire the world is abstracting how scientific thinking processes are practically applied to the examination of real-world problems. The present study considered that at the early stage of modeling an unknown learning mechanism in the world, analyzing data obtained from learning environments where causal relationships of the learning mechanism are relatively easy to identify (e.g., well-structured classroom learning) can be a first attempt to

 © ISLS

investigate a model that represents the basic structure of the learning mechanism and is potentially applicable to real-world settings. Following these ideas, for computationally representing the model in Fig. 1, this study analyzed scientific thinking processes included in lecture data in a classroom setting and extracted patterns of practical computations for formatively constructing internal representations of the world.

Analysis from a computational modeling approach

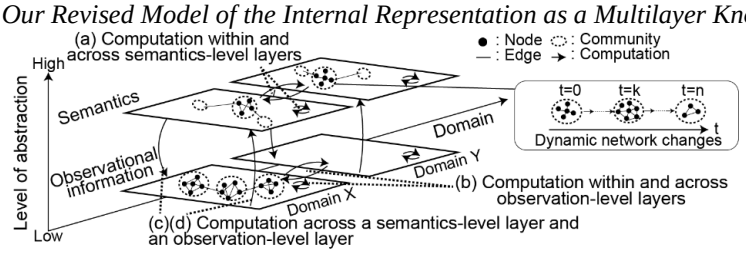
Based on the idea that a person's thinking process is externalized as his/her verbalization (Ericsson & Simon, 1980), the present study analyzed scientific thinking processes included in the data of utterances and slides of two lectures at universities, by using the data repository. The transcribed time-series utterance data were classified based on our initial assumptions of the information manipulations (a)–(d) in Fig. 1, from which practical computations were extracted. As systematic operations, this analysis extracted the following thinking processes, corresponding to (a)–(d) in Fig. 1, performed to encourage learners to scientifically acquire knowledge at the level of semantics in Fig. 1 (e.g., cross-disciplinary model representations):

- Establishing rational associations based on semantic similarities in real-world phenomena ((a) in Fig. 1).
- Filling in missing logic to provide possible explanations for the real-world phenomena ((b) in Fig. 1).
- Deductively interpreting observational information by using semantic knowledge ((c) in Fig. 1).
- Deriving semantic knowledge by induction of, analogy to, and generalization of observational information ((d) in Fig. 1).

Another finding from the viewpoint of the computational modeling approach is that each thinking process was able to be systematically broken down into some operational procedures. For example, thinking process (a) was broken into procedure (a1) utilizing axes and methodologies that uniformly explain different phenomena. This analysis suggested that these procedures were able to be regarded as practical computations for formatively constructing the internal representation. Furthermore, the present study considers that our initial model in Fig. 1 can be reformulated as a multilayer knowledge network with vertical layers representing levels of abstraction and horizontal layers representing domains such as disciplines (Fig. 2). Based on the graph representation of multilayer networks (Kivelä et al., 2014), our analysis enabled the model in Fig. 2 to be defined as $M = (V_M, E_M, V, \mathcal{L})$. V is a set of nodes. $\mathcal{L} = \{L_1, L_2\}$ is a sequence of sets of elementary layers $L_1 = \{Semantics, Observation\}$, $L_2 = \{DomainX, DomainY\}$. Each layer contains a set of node-layer tuples $V_M = \{(v, l_1, l_2) \mid v \in V, l_1 \in L_1, l_2 \in L_2\}$, which indicates which layer a node belongs to. E_M is a set of edges. The nodes form communities where nodes are densely connected. Importantly, one of the main achievements of the present study is that these definitions gave mathematical representations for several computations assumed in Fig. 2. For example, (a1) can be represented as an operation to apply $v_m = (v, l_1, l_2)$ to the other layers when v is a node expressing a certain axis or methodology. This study considers that these mathematical representations can be used to develop new metrics for assessing the process and results of learners' computations in Fig. 2. Our computational modeling approach is considered to be useful not only to enhance the traceability of learners' bottleneck processes in making scientific inquiries, but also to objectively describe what internal functions learners should optimize for better understanding of the world.

Figure 2

Our Revised Model of the Internal Representation as a Multilayer Knowledge Network



Conclusion

By examining how scientific thinking processes are performed to approach real-world problems, the present study analyzed the patterns of the information manipulations that learners should learn to formatively construct the internal representations of the world. The results of our analysis suggested that learners should learn information manipulations to acquire the cross-disciplinary model representations and these manipulations were able to be mathematically formulated as operational procedures to change the network structure of the multilayer internal representation. This study considers that our mathematical representations can be a basis for developing a CSCL system that computationally traces and assesses learners' scientific inquiries.

 © ISLS

References

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617-645.

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, *87*(3), 215-251.

Huys, Q. J. M., Browning, M., Paulus, M. P., & Frank, M. J. (2021). Advances in the computational understanding of mental illness. *Neuropsychopharmacology*, *46*, 3-19.

Kivelä, M., Arenas, A., Barthelemy, M., Gleeson, J. P., Moreno, Y., & Porter, M. A. (2014). Multilayer networks. *Journal of Complex Networks*, *2*(3), 203-271.

National Research Council. (2012). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. National Academies Press.

Watanabe, N., Tada, M., & Okada, M. (2023). Computational modeling of a learner's belief system underlying continuous real-world experiences. *In T. Bastiaens (Ed.), Proceedings of EdMedia + Innovate Learning*, 603–610. Vienna, Austria: Association for the Advancement of Computing in Education (AACE).

Acknowledgments

The authors thank Mr. Kazuma Tokunaga, who is a recent graduate of Kyushu University, Japan, for his support of our initial study. This study was funded by JSPS KAKENHI JP22K19797, JP23H03507, JP23K28197.

 © ISLS

International Society of the Learning Sciences

Orchestrating a Web-Based Classroom Game: Balancing Platform Automation and Teacher Work

Kai Yan, University of Toronto, kai.yan@alumni.utoronto.ca
Jasmine Hu, University of Waterloo, j37hu@uwaterloo.ca
James Slotta, University of Toronto, jim.slotta@utoronto.ca

Abstract: Multi group classroom games create a run time orchestration problem: teachers must coordinate pacing, content visibility, and group progress in real time. We examine this problem through a classroom enactment of Fall of Artica, a role-based inquiry game delivered in WISE. Using group centered audio, photographs, authoring screenshots, and planning notes, we analyzed 31 identifiable episodes across 13 step transitions, 14 post reveal intervals, and 4 whole class checkpoints. Findings show that lightweight continue gates and in place media enabled the platform to handle within step pacing and visibility, while teacher intervention clustered around logistics recovery, brief procedural alignment, and concept advancing prompts. Whole class convergence remained teacher led, and externally embedded tools increased local recovery work. These findings suggest that, in role-based classroom inquiry, automation is more effective for pre-structured sequencing and visibility than for interpretive, contingent, and convergence work. We argue that platform support redistributed rather than replaced teacher orchestration, and we derive three design implications for the run time enactment of classroom games.

Introduction

In multi group classroom games, teachers must coordinate activity flow, timing, and shared attention across several groups at once. Prior CSCL work has described orchestration as the coordination of tools, time, and social interaction in technology rich learning environments (Dillenbourg et al., 2011; Prieto et al., 2015; Roschelle et al., 2013). More recent work has moved beyond defining orchestration as a general classroom challenge and has begun to examine more specifically how orchestration load is distributed across forms of teacher support and technology provision (Amarasinghe et al., 2021; Hakami et al., 2024). As platforms increasingly support sequencing and content delivery, an important question is not simply whether automation helps, but which orchestration tasks become manageable by the platform and which remain teacher dependent during class.

We examine this question in a classroom enactment of Fall of Artica (FoA), a role-based inquiry game run in the web-based Inquiry Science Environment (WISE) (Linn et al., 2003; Slotta & Linn, 2009). Recent work on critical action games and related role-based inquiry environments has explored how dystopian scenarios, role structures, and agent-based supports can scaffold collective inquiry and participation (Zhou et al., 2023; Zhou, 2024). Students in our enactment adopted roles and pursued region-specific quests, while WISE delivered instructions, media, and step-based activity flow. We focus on run time orchestration, meaning what happened during class as groups moved through steps, encountered new content, and periodically reconverged. Our questions were: (1) Which orchestration tasks were handled by the platform during class, and which remained teacher work? (2) Across key classroom episodes, when and how did the teacher intervene?

Context and methods

The study took place in an undergraduate critical literacies course. Students worked in four table groups through four FoA regional units. WISE provided self-paced step navigation, one continue gate that revealed follow up content after a prerequisite task, in place media, and one embedded chatbot activity.

Data sources included group centered table audio, photographs, authoring screenshots, and planning notes.

Episodes were the unit of analysis. We analyzed all identifiable episodes meeting our inclusion criteria in the focal enactment, including 13 step transition episodes, 14 post reveal episodes, and 4 whole class checkpoint episodes. A step transition episode was included when movement into a new step could be identified through teacher cueing, student talk, or corroborating visual artifacts such as photographs or authoring screenshots. A post reveal episode was included when a continue action and the subsequent one-to-two-minute interval could be confirmed through synchronized audio, screenshots, or other classroom records. A whole class checkpoint episode was included when the instructor explicitly announced a moment of class-wide convergence.

For each episode, we identified which platform supports were active at run time and classified teacher moves as logistics, procedural alignment, or conceptual prompting. Triangulation across audio, screenshots, photographs, and planning notes was used to confirm episode timing, content visibility, and classroom alignment.

 © ISLS

Because silent self-paced transitions were not always observable in the available data, the dataset should be understood as including all identifiable episodes meeting our criteria rather than all possible transitions that may have occurred during the lesson.

Findings

Across episodes, the platform most reliably handled within step pacing and visibility. Continue gates staged the sequence of actions and reduced repeated teacher reminders about when to move forward. In place images and short media also supported smooth flow because students could access relevant content without switching tools or files.

Teacher work did not disappear. Instead, it shifted toward contingent local support. During step transitions, teacher intervention was usually brief and procedural, helping lagging groups identify the next action or rejoin the current step. During post reveal intervals, teacher talk more often became conceptual, pressing students to interpret new material, connect it to prior notes, or clarify what would count as evidence. These patterns suggest that while the platform could structure routine sequencing, it could not absorb the interpretive work needed to advance inquiry.

Whole class convergence remained teacher led. Although WISE supported group level progression, it did not synchronize stopping points across groups. During the tavern checkpoint, the teacher announced the start and end of the exchange, monitored lagging groups, and provided brief support so all students could participate. External embeds also revealed a boundary condition: the chatbot added useful interactional possibilities, but it introduced startup and recovery demands that increased local logistics support.

Taken together, these patterns suggest that platform support was most effective for pre-structured sequencing and visibility, while teacher intervention remained necessary in moments requiring local judgment, interpretation, and class-wide realignment.

Discussion and design implications

The main contribution of this case is not simply that teachers still help during classroom games. Rather, it shows a more specific division of labor under partial automation. Tasks that can be pre-structured in advance, such as within step sequencing and content visibility, are more readily absorbed by the platform. Tasks that require contingent judgment, local recovery, conceptual prompting, or whole class realignment remain teacher dependent. In this sense, automation redistributed rather than replaced teacher orchestration.

This interpretation aligns with recent work on orchestration load, which has argued that technology can reduce some aspects of classroom coordination while leaving teachers responsible for managing instructional and social complexity in real time (Amarasinghe et al., 2021; Hakami et al., 2024). More specifically, our findings suggest that within step sequencing and visibility were more readily absorbed by the platform, whereas conceptual prompting, local recovery, and whole class convergence remained teacher dependent. Relative to recent work on critical action games and role-based inquiry settings (Zhou et al., 2023; Zhou, 2024), the contribution here is not simply to argue that such designs can support engagement or inquiry, but to show how they are orchestrated moment by moment across multiple groups during classroom enactment. In this sense, the paper contributes a more fine-grained account of how automation redistributes teacher work in role-based classroom inquiry rather than simply making classroom games easier to run.

First, use minimal gates to stage pacing and visibility. Continue gates reduced routine sequencing talk after reveals and helped groups progress with less teacher prompting. Second, keep media in place and embed externally only when clearly necessary. Inline media supported smoother enactment, while external tools introduced startup and recovery work. Third, reserve teacher effort for concept support and brief alignment. Teacher intervention was most valuable when focused on helping students interpret new material, reconnect to the task, or rejoin shared classroom moments.

Conclusion

In this enactment of Fall of Artica, the platform reliably handled within step pacing and content visibility, while teacher intervention clustered around logistics recovery, brief procedural alignment, and concept advancing prompts. Whole class convergence remained teacher led. For CSCL, the significance of this case lies in showing that platform support reorganizes rather than removes teacher orchestration. Effective classroom game design may therefore depend less on maximizing automation than on combining lightweight platform structure with strategically timed teacher intervention.

 © ISLS

References

Amarasinghe, I., Hernández-Leo, D., & Hoppe, H. U. (2021). Deconstructing orchestration load: comparing teacher support through mirroring and guiding. International Journal of Computer-Supported Collaborative Learning, 16(3), 307-338. https://doi.org/10.1007/s11412-021-09351-9

Dillenbourg, P., Zufferey, G., Alavi, H., Jermann, P., Do-Lenh, S., Bonnard, Q., ... & Kaplan, F. (2011). Classroom orchestration: The third circle of usability. https://doi.org/10.22318/cscl2011.510

Hakami, L., Hernández-Leo, D., Amarasinghe, I., & Sayis, B. (2024). Investigating teacher orchestration load in scripted CSCL: A multimodal data analysis perspective. British Journal of Educational Technology, 55(5), 1926–1949. https://doi.org/10.1111/bjet.13500

Linn, M. C., Clark, D., & Slotta, J. D. (2003). WISE design for knowledge integration. Science Education, 87(4), 517-538. https://doi.org/10.1002/sce.10086

Prieto, L. P., Dimitriadis, Y., Asensio-Pérez, J. I., & Looi, C.-K. (2015). Orchestration in learning technology research: Evaluation of a conceptual framework. Research in Learning Technology, 23, 28019. https://doi.org/10.3402/rlt.v23.28019

Roschelle, J., Dimitriadis, Y., & Hoppe, U. (2013). Classroom orchestration: synthesis. Computers & Education, 69, 523-526. https://doi.org/10.1016/j.compedu.2013.04.010

Slotta, J. D., & Linn, M. C. (2009). WISE science: Web-based inquiry in the classroom. Teachers College Press.

Zhou, K. H., Pullen, C., Holmes, J., & Slotta, J. D. (2023). Supporting collective inquiry in a critical action game: A role for open AI conversational agents. In Proceedings of the 16th International Conference on Computer-Supported Collaborative Learning-CSCL 2023, pp. 428-431. International Society of the Learning Sciences. https://doi.org/10.22318/cscl2023.922487

Zhou, K. H., Pullen, C., & Slotta, J. D. (2024). Digging the Abandoned Feelings out of the Wasteland: Raising Trouble Makers through A Dystopian Critical Action Game. In *Proceedings of the 18th International Conference of the Learning Sciences – ICLS 2024* (pp. 2509–2510).

International Society of the Learning Sciences

Meet Bip: Understanding Student–Peer Agent Interaction in an Embodied Exploratory Learning Environment

Erfan Farhadi, Kenneth Fei, Yifan Jiang, Zhen Bai
efarhadi@cs.rochester.edu, kfei@u.rochester.edu, yjiang54@ur.rochester.edu, zhen.bai@rochester.edu
Department of Computer Science, University of Rochester, Rochester, NY, USA

Abstract: Peer agents have been shown to enhance learning in K–12 settings, yet the design of LLM-powered embodied peer agents for exploratory learning remains underexplored. This study examines how middle-school students interact with Bip, an embodied conversational agent in a Minecraft-based AI literacy game. Inspired by the Intelligent Novice paradigm, Bip observes players, provides hints, and scaffolds reflection during an AI literacy activity. Findings from nine students highlight the importance of spatial awareness, responsiveness, and bounded understanding in supporting exploratory learning. We discuss design implications for embodied agents that foster scientific inquiry.

Introduction

Peer learning, where learners collaborate and reason as equals, is a powerful driver of curiosity, reflection, and scientific inquiry (Cassell, 2022; Paranjape, Ge, et al., 2018). Peer agents extend this idea into interactive systems by modeling social and cognitive aspects of collaboration, ranging from chat-based agents to embodied conversational agents (ECAs) in virtual environments. At the same time, exploratory learning environments (ELEs) support inquiry through experimentation, hypothesis testing, and observation (Alawajee & Delafield-Butt, 2021; Y. Zhou & Paquette, 2024). Recent advances in large language models (LLMs) further enable agents to sustain context-rich dialogue and scaffold reflection, making conversational peer agents increasingly effective in supporting learning (Earle-Randell et al., 2024; Liu et al., 2024). However, despite progress in both embodied agents and LLM-based conversational systems, little work has integrated the two to create LLM-powered embodied peer agents that reason and interact through both dialogue and spatial action in ELEs. In this work, we build on prior research in embodied AI literacy (Farhadi et al., 2026; Zhou et al., 2024) to study how middle-school students interact with such an agent in a Minecraft-based environment. Inspired by the Intelligent Novice framework (Mathan & Koedinger, 2005), our agent is designed to observe, question, and respond to learners' actions rather than provide direct instruction. Through qualitative analysis of student–agent interactions, we examine how these agents can support exploration, reflection, and inquiry.

Meet Bip's system design

Bip is an ECA implemented as a bee within a Minecraft-based AI literacy game, where players learn about filter bubbles in recommendation systems. The agent processes player actions, spoken input, and in-game events to provide context-sensitive hints and scaffold reflection. Bip exhibits both verbal and nonverbal cues, including gaze and proximity; while facial expressions are absent, dialogue is displayed above the agent in synchrony with audio (Figure 1). Guided by the Intelligent Novice framework, Bip is designed to demonstrate partial understanding, encouraging peer-like interaction. Bip integrates GPT-4o for dialogue and decision-making, Fabric for Minecraft modding, Whisper for speech recognition, and Typecast for voice synthesis.

Figure 1

Screenshots of Bip Introducing Itself, and Reacting to the Learner Pollinating a Flower



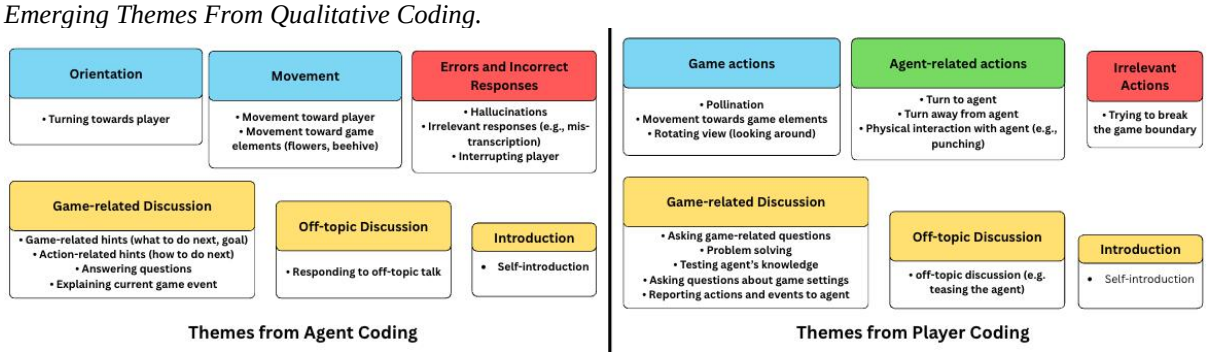
 © ISLS

Methods

Participants consisted of nine middle-school students (grades 6–9; seven male, two female), predominantly from African American backgrounds. These learners possessed high Minecraft experience and moderate AI familiarity, with limited prior programming knowledge. During a summer-camp study, students worked in pairs to complete the 30–40 minute Meet Bip activity. All sessions were recorded (capturing screen, audio, and video) and subsequently transcribed using the Whisper model for qualitative analysis. Consent and assent were obtained in accordance with ethical guidelines. We conducted an inductive thematic analysis (Blandford et al., 2016; Braun & Clarke, 2013; Zhou et al., 2025), iteratively coding salient interactions (e.g., dialogue, movement, engagement with game elements) and clustering them into higher-level categories capturing patterns in Bip's behavior and students' responses. Two researchers independently developed the codebook from session recordings and refined codes through cross-session comparison to ensure consistency.

Figure 2

Emerging Themes From Qualitative Coding.



Yellow blocks represent verbal behaviors, blue blocks represent nonverbal behaviors, green blocks indicate interactions with the agent, and red blocks denote errors or boundary-violating behaviors.

Results and discussion

To analyze student–agent interaction, we coded Bip's and players' behaviors across verbal and nonverbal dimensions (Figure 2). Figure 2 summarizes the coded behavioral categories for both Bip and players; by examining how these behaviors co-occurred across interactions, we identified recurring interaction patterns that informed the synthesis of three design dimensions: spatial understanding, responsiveness, and bounded knowledge. Bip generally maintained coherent, contextually relevant dialogue, while these three factors proved most influential in shaping interaction quality.

When Bip grounded its utterances in the spatial context, such as naming nearby flowers or moving toward relevant elements, players displayed higher engagement and moments of shared attention. Conversely, missing contextual details prompted players to use "repair talk" to re-establish common ground, repositioning Bip as an NPC requiring updates rather than a peer participant. Responsiveness was also critical; while smooth exchanges supported reasoning, latency from transcription or LLM processing led to frustration and disengagement.

Interestingly, Bip's intentional limitations served a constructive function. Small factual or interpretive errors encouraged students to articulate reasoning and engage in productive friction, such as argumentative exchanges that reinforced understanding. This suggests that bounded imperfection can be pedagogically valuable, as it positions the agent as an "intelligent novice" rather than an infallible guide. However, deep integration in an exploratory learning environment (ELE) requires both spatially coherent referencing and low-latency responsiveness. When these slipped, players defaulted to NPC expectations rather than treating Bip as a partner in exploration.

These findings highlight that agents must be perceptually aware and capable of joint attention to establish presence. Responsiveness sustains interaction flow, while limited knowledge prompts learners to question and reason. By acting as an intelligent novice, Bip redirected attention to overlooked elements and posed interpretive questions, scaffolding reflection while maintaining exploration. Future development should focus on richer multimodal tracking and adaptive response timing to investigate how embodied agents can dynamically balance curiosity and autonomy. While the findings are based on a small, exploratory sample (n = 9) and should be interpreted as indicative rather than generalizable, even with a constrained prototype, this study demonstrates that responsiveness, spatial grounding, and bounded knowledge collectively shape whether learners treat an agent as a co-participant or a simple NPC guide. As conversational models evolve, the challenge is to design inquisitive partners that prompt discovery through interaction.

References

Aljammaz, R., Wardrip-Fruin, N., & Mateas, M. (2023). Towards an Understanding of Character Believability. *Proceedings of the 18th International Conference on the Foundations of Digital Games*, 1–9. https://doi.org/10.1145/3582437.3582466

Blandford, A., Furniss, D., & Makri, S. (2016). *Qualitative Hci Research: Going Behind the Scenes*. Morgan & Claypool Publishers.

Bonetti, F., Bucchiarone, A., & Wanick, V. (2025). Using LLMs to Adapt Serious Games with Educators in the Loop. In A. Schönbohm, F. Bellotti, A. Bucchiarone, F. de Rosa, M. Ninaus, A. Wang, V. Wanick, & P. Dondio (Eds.), *Games and Learning Alliance* (pp. 68–77). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-78269-5_7

Braun, V., & Clarke, V. (2013). *Successful qualitative research: A practical guide for beginners* (First published). SAGE.

Cassell, J. (2022). Socially Interactive Agents as Peers. In B. Lugrin, C. Pelachaud, & D. Traum (Eds.), *The Handbook on Socially Interactive Agents* (1st ed., pp. 331–366). ACM. https://doi.org/10.1145/3563659.3563670

Chatterji, A., Cunningham, T., Deming, D., Hitzig, Z., Ong, C., Shan, C., & Wadman, K. (n.d.). *How People Use ChatGPT*.

Chen, H., Park, H. W., & Breazeal, C. (2020). Teaching and learning with children: Impact of reciprocal peer learning with a social robot on children's learning and emotive engagement. *Computers & Education*, *150*, 103836. https://doi.org/10.1016/j.compedu.2020.103836

Earle-Randell, T. V., Wiggins, J. B., Ma, Y., Celepkolu, M., Bounajim, D., Gao, Z., Ruiz, J. M., Boyer, K. E., Israel, M., Lynch, C. F., & Wiebe, E. (2024). The impact of near-peer virtual agents on computer science attitudes and collaborative dialogue. *International Journal of Child-Computer Interaction*, *40*, 100646. https://doi.org/10.1016/j.ijcci.2024.100646

Farhadi, E., Fei, K., Jiang, Y., & Bai, Z. (2026). Beetrap-MC: A Minecraft-Based AI Literacy Tool for Teaching Filter Bubbles to Middle School Students. *Proceedings of the AAAI Conference on Artificial Intelligence*, *40*(47), 40754–40761. https://doi.org/10.1609/aaai.v40i47.41524

Gallotta, R., Todd, G., Zammit, M., Earle, S., Liapis, A., Togelius, J., & Yannakakis, G. N. (2024). Large Language Models and Games: A Survey and Roadmap. *IEEE Transactions on Games*, 1–18. https://doi.org/10.1109/TG.2024.3461510

Kim, Y., & Baylor, A. L. (2016). Research-Based Design of Pedagogical Agent Roles: A Review, Progress, and Recommendations. *International Journal of Artificial Intelligence in Education*, *26*(1), 160–169. https://doi.org/10.1007/s40593-015-0055-y

Liu, T., Zhao, H., Liu, Y., Wang, X., & Peng, Z. (2024). ComPeer: A Generative Conversational Agent for Proactive Peer Support. *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*, 1–22. https://doi.org/10.1145/3654777.3676430

Mathan, S. A., & Koedinger, K. R. (2005). Fostering the intelligent novice: Learning from errors with metacognitive tutoring. *Educational Psychologist*, *40*(4), 257–265. https://doi.org/10.1207/s15326985ep4004_7

Paranjape, B., Bai, Z., & Cassell, J. (2018). Predicting the Temporal and Social Dynamics of Curiosity in Small Group Learning. In C. Penstein Rosé, R. Martínez-Maldonado, H. U. Hoppe, R. Luckin, M. Mavrikis, K. Porayska-Pomsta, B. McLaren, & B. Du Boulay (Eds.), *Artificial Intelligence in Education* (Vol. 10947, pp. 420–435). Springer International Publishing. https://doi.org/10.1007/978-3-319-93843-1_31

Paranjape, B., Ge, Y., Bai, Z., Hammer, J., & Cassell, J. (2018). Towards Automatic Generation of Peer-Targeted Science Talk in Curiosity-Evoking Virtual Agent. *Proceedings of the 18th International Conference on Intelligent Virtual Agents*, 71–78. https://doi.org/10.1145/3267851.3267894

Starks, K. (2014). Cognitive behavioral game design: A unified model for designing serious games. *Frontiers in Psychology*, *5*, 28. https://doi.org/10.3389/fpsyg.2014.00028

Acknowledgments

This work is supported by the National Science Foundation (NSF) RETTL program (Award No. 2238675).

We would like to thank Jeremy Smith, Kelvin Sanchez-Orona, and the Servant leaders and Junior Servant leaders at the Freedom Scholars Learning Center for their generous support, as well as scholars who participated in our study and their parents.

International Society of the Learning Sciences

Dropout and Success in Online Higher Education: A Comparative Analysis of Students in Brazil

Andréia Maria de Lima Assunção, University of São Paulo, amlassuncao@gmail.com
Paula Daniela Zapata, University of São Paulo, zapata.pauladaniela@gmail.com
Lívia Pizauro Sanchez, University of São Paulo, livia.pizauro@gmail.com

Abstract: Despite the significant growth of online higher education, challenges persist regarding how to foster engagement and address dropout rates. Therefore, personalized learning has been identified as bridges to success. This paper examines whether the design of reasonable accommodations based on personalized learning affects the success or dropout rates of students in undergraduate courses at an online higher education university in Brazil. By tabulating and analyzing the public database of student enrollment status before and after the design changes, we observed that personalized learning was associated with higher success rates and lower dropout rates, suggesting that Universal Design for Learning (UDL) and Personalized Learning (PL) can mutually benefit a diverse student population.

Introduction and theoretical framework

In many Global South nations, democratizing access to higher education has become a political priority, with online higher education (OHEs) institutions increasingly seen as a scalable way to expand enrollment (Flores et al., 2010; Chiroleu & Marquina, 2017). However, these institutions continue to face significant challenges, particularly high dropout rates in large-scale distance learning programs (Shaik & Asif, 2022; Yang et al., 2025) and, consequently, low graduation rates (Smaili et al., 2022). To address these issues, researchers and educators have explored personalized design approaches (Lin et al., 2024) to prevent at-risk students from dropping out (Smaili et al., 2022) and to improve educational outcomes (Heiman, 2006). In this context, Universal Design for Learning (UDL) and Personalized Learning (PL) frameworks have been applied to create environments that accommodate variability and provide multiple forms of engagement, gaining wide policy adoption in equity-focused education (Meyer et al., 2014; Zhang et al., 2020). The integration of these modalities offers a strategy that benefits students with and without Specific Educational Needs (SEN) (Zhang et al., 2022), shifting focus from individual deficits to systemic barriers to learning (Sewell et al., 2022). This perspective aligns with the Brazilian concept of SEN, which defines the target group for specialized educational services and guides the design of resources that remove barriers to student participation (Mantoan, 2003).

Building on this, the study examines success and dropout rates among undergraduates at a public, tuition-free OHE university in Brazil. Using data from the public database covering 2014-2023, we assessed whether Personalized Learning (PL) interventions introduced in 2018 influenced outcomes for students with and without SEN. With this, we ask: RQ1: What are the success and dropout rates of students with (G1) and without (G2) SEN? RQ2: Did the adoption of PL-based design affect the success and dropout rates of students with SEN (G2) before and after 2018? Enrollment data from students, combined with historical events at the university, were analyzed to explain disparities. Results show significant changes after PL implementation for both groups, especially for non-SEN students, who experienced a marked reduction in dropout rates alongside an increase in success rates. Nevertheless, the dropout levels observed warrant further specific studies, so that structural inequalities that continue to constrain equitable access, retention, and graduation in OHE institutions in Brazil can be more effectively addressed and contextualized.

Methods

Context, collection and data analysis

The university, founded in Brazil in 2014 as a public and tuition-free OHE institution, has ensured since its inception that its Virtual Learning Environment (VLE) and test system comply with WCAG standards. Beginning in 2018, increasing accommodation requests from SEN students prompted system redesigns and ultimately led to the creation of an institutional committee dedicated to accessibility and inclusion. Data were obtained from the university's public dashboard, which provides annual institutional information covering the period from 2014 to 2023. The enrollment status of 134,501 students was tabulated and categorized into two groups: G1 (N = 1,582), referring to students with SEN, and G2 (N = 132,919), referring to students without SEN. Subsequently, the data

 © ISLS

were proportionally compared using a two-proportion z-test. Next, applying two-proportion z-test, we compared data from both groups before and after the implementation of reasonable accommodations grounded in personalized learning, introduced in response to increasing student diversity following the expansion of university access. The personalized learning strategies requested by students and adopted by the university included granting extended exam time, early release of content, conducting remote exams (instead of in-person exams at the university), developing activities individually that would originally be collective, and improving captioning, sign language interpretation, audio description, and compatibility of the entire system with screen reader software. In this view, the interpretive process was interactive and involved correlating the findings with the institution's historical context.

Findings

A first analysis was conducted on the existing data regarding students with SEN. Between 2014 and 2023, a total of 1,582 enrolments were recorded. The years 2018 and 2022 registered the highest number of SEN students, whereas the lowest numbers were observed between 2014 and 2016. Overall, 898 students were classified as having achieved academic success, while 684 were unable to continue their studies. In the second phase of the analysis, the academic progression of 132,919 non-SEN students was examined. Mirroring the trend observed among SEN students, the highest enrolment numbers were recorded in the years 2018 and 2022, while the lowest occurred in 2014, 2015, and 2016. Of the total non-SEN student population, 60,026 were classified as academically successful, whereas 72,893 were unable to complete their studies. Following the individual analysis of both groups, a proportional comparison of success and failure rates was conducted. A two-proportion z-test showed statistically significant difference in enrollment success rates between SEN and non-SEN students ($p = 0,0001$), demonstrating distinct academic performance across the two groups.

Discussion and conclusions

In Brazil, several policies were implemented to democratize access to public higher education, leading to an increase in enrollment from 6.7 million in 2011 to nearly 8.9 million students in 2021 (Brasil, 2021). The online university examined in this study was established in the early 2010s, although these policies could only be effectively implemented beginning in 2018. This timeline helps explain the rise in the number of applications and, consequently, the increase in students with SEN.

Considering RQ1 (What are the success and dropout rates of students with (G1) and without (G2) SEN?), the data show that 898 out of 1,582 SEN students with SEN achieved academic success (approximately 56,76%), while 684 (approximately 43,23%) dropped out. Among non-SEN students, 60,026 out of 132,919 (approximately 45,16%) were successful, and 72,893 (approximately 54,84%) did not complete their studies.

Regarding RQ2 (Did the adoption of PL-based design affect the success and dropout rates of students with SEN (G2) before and after 2018?), our findings indicate that success rates in both groups increased. Following the adoption of PL-based design, the success rates of SEN students (G2) rose from 51.75% to 58.46%, representing an increase of 6.71%. In contrast, the success rates of non-SEN students (G1) improved much more significantly, increasing from 29.28% to 54.74%, a gain of 25.46%. Before the adoption of PL-based design, the success rate of G1 was 22.47% lower than that of G2. However, after the implementation of the changes, this difference fell to 3.72%. These findings reinforce the results reported by Zhang et al. (2022), who predict that Universal Design for Learning (UDL) and Personalized Learning (PL) mutually benefit students with and without SEN.

In turn, dropout rates decreased from 48.25% to 41.54% for G2 (SEN) and showed a sharp reduction from 70.72% to 45.26% for G2 (non-SEN) following the implementation of the PL-based design approach. Despite these improvements, the rates remain high. We therefore emphasize the need for further studies that analyze the enrollment cancellation forms completed by students themselves to identify the potential reasons associated with the high failure rates observed, so that indicators for addressing them within this university can be delineated. In addition, future research should consider correlations with factors that may contribute to dropout in a post-pandemic period (Kaisara et al., 2024)

We conclude that inclusive pedagogy, grounded in personalized learning, is inherently nuanced and context-dependent. At this university, such resources have shown strong potential in improving student success rates and reducing dropout rates across different groups. However, current data do not yet allow us to establish a direct causal link between the observed outcomes and the PL measures adopted, underscoring the need for more targeted studies. In this regard, it is important to emphasize that in regions of the Global South, where structural inequalities persist, this approach must be complemented by broader systemic support to effectively address barriers to equitable access, retention, and graduation in online higher education, particularly given the significant dropout rates observed.

References

Brasil. (2021). *Censo da Educação Superior 2021: Divulgação dos resultados*. Ministério da Educação.

Chiroleu, A., & Marquina, M. (2017). Democratization or credentialism? Public policies of expansion of higher education in Latin America. *Policy Reviews in Higher Education*, 1(2), 139–160. https://doi.org/10.1080/23322969.2017.1303787

Flores, G., Monroy, G. V., García, C. F., & Sánchez, E. N. (2010). *Políticas compensatorias para la educación superior en 10 países de América Latina*. Unión de Universidades de América Latina y el Caribe and Universidad Autónoma del Estado de México.

Heiman, T. (2006). Assessing Learning Styles among Students with and without Learning Disabilities at a Distance-Learning University. *Learning Disability Quarterly*, 29(1), 55–63. https://doi.org/10.2307/30035532

Kaisara, G., Peel, C., Niemand, C., & Bwalya, K. J. (2024). Exploring factors influencing e-learning dropout rates in the post-COVID-19 era. *International Journal of Information and Communication Technology Education*, 20(1). https://doi.org/10.4018/IJICTE.348660

Lin, L., Lin, X., Zhang, X., & Ginns, P. (2024). The Personalized Learning by Interest Effect on Interest, Cognitive Load, Retention, and Transfer: A Meta-Analysis. *Educ Psychol Rev*, 36(88). https://doi.org/10.1007/s10648-024-09933-7

Meyer, A., Rose, D. H., & Gordon, D. (2014). Universal design for learning: Theory and Practice. CAST Professional Publishing.

Mantoan, M. T. E. (2003). *Inclusão escolar: o que é? por quê? como fazer?*. Moderna.

Sewell, A., Kennett, A., & Pugh, V. (2022). Universal Design for Learning as a theory of inclusive practice for use by educational psychologists. *Educational Psychology in Practice*, 38(4), 364–378. https://doi.org/10.1080/02667363.2022.2111677

Smaili, E. M., Khoudda, C., Sraidi, S., Azzouzi, S., & Charaf, M. E. H. (2022). An Innovative Approach to Prevent Learners' Dropout from MOOCs using Optimal Personalized Learning Paths: An Online Learning Case Study. *Statistics, Optimization & Information Computing*, 10(1), 45-58. https://doi.org/10.19139/soic-2310-5070-1206

Yang, L., Ziyang, S., & Peng, W. (2025). A systematic analysis of the problem of MOOC learner dropouts. In *Proceedings of the 2025 International Conference on Big Data and Informatization Education*. (pp. 266–271).

Zhang, L., Basham, J. D. and Yang, S., 2020. Understanding the implementation of personalized learning: A research synthesis. *Educational Research Review*, 31, 100339. https://doi.org/10.1016/j.edurev.2020.100339

Zhang, L., Carter, R. A. Jr, Basham, J. D., & Yang, S. (2022). Integrating instructional designs of personalized learning through the lens of universal design for learning. *Journal of Computer Assisted Learning*, 38(6), 1639–1656. https://doi.org/10.1111/jcal.12725

Acknowledgments

The authors acknowledge the support of the Coordination for the Improvement of Higher Education Personnel (CAPES) and the National Council for Scientific and Technological Development (CNPq), both research funding agencies linked to the Ministry of Education in Brazil, in their training as researchers.

International Society of the Learning Sciences

Author Index

A. Susan Jurow, 3
Adiel Benjamin, 484, 577
Ahmad Khanlari, 139, 439, 484, 752
Akash K. Saini, 454
Akin Adediji, 728
Alae Sherafat, 761
Alejandra J Magana, 129
Alexandra Gendreau Chakarov, 632
Ali Keramati, 770
Alina Guha, 274
Alina Sheppard-Bujtor, 794
Alison Nelson, 621
Alyssa Wise, 566, 589
Amanda M. Cottone, 674
Amy Ko, 148
Anders Mørch, 61
Andréia Maria de Lima Assunção, 812
Andrés Felipe Zambrano, 489
Ane Vielandt Jensen, 359
Anna Pirner, 414
Annamarie Brosnihan, 414, 424
Anouschka van Leeuwen, 399
Anthony F. Botelho, 244, 489, 767
Anupam Raj, 549
Aparajita Jaiswal, 615
Aric Rindfleisch, 283
Ariel Han, 364, 644
Armando Paulino Preciado Babb, 621
Arooj Zaidi, 779
Asad Shahab, 224
Ashish Ashish, 615
Ashwin T S, 194
Avery Closser, 244
Aya Isaac, 429
Ayano Ohsaki, 788
Ban Mouid Shiwalia, 566
Batuhan Sayis, 692
Bo Shui, 731
Bo Yu, 641
Bodong Chen, 484, 577
Boris Streipe, 484
Breneil Malcolm, 566
Brian Magerko, 683
Brooklyn Clines, 556
Bunichi Otaki, 677
C. Naomie Williams, 800
Caitlin Morris, 175
Cameron Marano, 414, 424
Can Sun, 42
Carl Bereiter, 484, 577
Carol K.K. Chan, 514, 577
Carolina Sotério, 743
Caroline Traverse, 484
Carolyn Rose, 3, 224, 758
Casey Smith, 650
Catherine Roeroe, 519
ChanMin Kim, 653
Charis Charitsis, 746
Charlotte Beal, 234
Chelsea Chandler, 556
Chen Feng, 354
Chen Li, 148
Chen Wang, 577, 737
Chencheng Le, 70
Chenxuan Meng, 785
Chenyu Hou, 624
Chew Lee Teo, 577
Chih Hui Seet, 409
Chih-Hsuan Chang, 409
Chihiro Sato, 779
Ching Sing Chai, 524
Ching-Hui Kuan, 539
Chloe Lok, 479
Chris Palaguachi, 494
Christiam Ivan Mendoza García, 419
Christian Dindler, 359
Christiane Lütge, 698
Christopher Hughes, 454
Chunlin LEI, 782
Cindy E. Hmelo-Silver, 354, 504, 647
Claudia Susana López Cruz, 419
Clayton Cohn, 509, 647
Collette Heskett, 369, 556
Connor Bain, 185
Daeun Hong, 354, 504
Daisy Fernandez-Reyes, 644
Daner Sun, 254, 264, 773
Daniel Tsivkovski, 364
Danielle P. Espino, 309
Davinia Hernández-Leo, 99, 155, 299, 589, 692, 707
Dejan Trencevski, 283
Dengkang Chen, 389
Di Fan, 734
Dina Soliman, 577
Dongni Guo, 761
Donna Dancer, 719
Dorentina Dedushaj, 749
Drew Gossen, 764
Dylan Paré, 656, 686
Dylan Ravel, 364
Ean Teng Khor, 344
Eda Zhang, 618
Eduard Arnau Romeu, 692
Ekta Shokeen, 359
Elaine Khoo, 484
Eliane Chaker Metni, 484
Elizabeth Koh, 344
Emily Kerzabi, 379, 668

 © ISLS

Emily Watts, 556
Emmanuel Dorley, 494
Engin Bumbacher, 662
Enruo Guo, 379, 668
Erfan Farhadi, 224, 809
Erfane Ghasempour Khoshroodi, 534
Eric R. Hamilton, 309
Erika Cheek, 659
Fabio Campos, 743
Fanjie Li, 589
Franceli Cibrian, 364, 644
Frank Fischer, 722
Freydis Vogel, 504, 698
Fucheng Guo, 389
Fuxin KONG, 782
Gahyun Callie Sung, 734
Gaowei Chen, 32, 42, 524, 695
Gaoxia Zhu, 70, 254, 624
Gautam Biswas, 194, 509, 647
Giulia Barbareschi, 779
Gordon Yeung, 746
Grace Lin, 749
Gregory Benedis-Grab, 556, 632
Greta Heathcote, 621
Griselda Esthela Oyervides Ramírez, 419
Guangji Yuan, 577
Guo Su, 70
Gur Arie Livni Alcasid, 725
Gustavo Oliveira, 743
H. Chad Lane, 494
Haiyang Xin, 524, 710
Haiyi Zhu, 800
Hakeoung Hannah Lee, 680
Hangfeng He, 224
Hans C. Arnseth, 589
Hanxiang Du, 624
Haowei Li, 674
Haozheng Du, 224
Hecong Wang, 224
Heliya Ahmadi, 656, 686
Henrik Tjønn, 61
Hideki Kondo, 716
Hideo Funaoi, 704
Hideya Matsukawa, 788
Hideyuki Suzuki, 704
Hironori Egi, 214, 665
Hiroshi Kato, 704
Ho Sue Kwon, 519, 671
Hongming Li, 244, 489, 767
Hongwen Guo, 148
Huang-Yao Hong, 409
Hyejeong Lee, 719
Ibrahim Oluwajoba Adisa, 204, 499
Ilana Dubovi, 725
Imogen Casebourne, 304, 641
Indrani Dey, 369, 653, 689
Inés Álvarez-Icaza, 419
Ingo Kollar, 80, 794
Injila Rasul, 474
Iris Saban, 294
Iris Tabak, 725
Isabella Ruel, 749
Ishari Amarasinghe, 155
Ishita Asnani, 731
Ivon Arroyo, 474
Jacob Davidsen, 234
Jacob Hale, 755
Jacob Puthipiroj, 797
Jacqueline Tan, 624
Jaeyoon Choi, 374
James D. Slotta, 534, 785, 806
James Lester, 354, 647
Jasmine Hu, 806
Jaycee Sansom, 713
Jeffrey Bush Tayne, 556, 689
Jenna Matthews, 15
Jennie Lee, 494
Jennifer Kahn, 51
Jennifer Muirhead, 761
Jennifer Osuna, 746
Jerika Miller, 504
Jessica Andrews-Todd, 379, 668
Ji-Eun Lee, 244, 489, 767
Jiangang Hao, 379, 668
Jianhua Zhao, 514
Jianwei Zhang, 429, 577, 761
Jianwen Sun, 264, 389
Jiayu Cheng, 577
Jina Kang, 283, 449, 494
Jingjing Zhang, 6
Jinhee Lee, 728
Jiyoung Cheong, 624
Jo Inge J. Froytlog, 589
Joanne Barrett, 494
Joey Huang, 758
John Mitchell, 746
John Weatherley, 556
Jon M. Wargo, 459
Jooeun Shim, 671
Joshua Danish, 647
Joyce Fonteles, 194
Juan Carlos Márquez Cañizares, 419
Juan Zheng, 624
Juhong Eom, 734
Julia Eden, 556
Julia Hornstein, 80, 794
Jun Lu, 638
Jun Oshima, 329, 334, 484, 577, 638, 677
Junichi Yamaoka, 779
Kai Yan, 806
Karina Vázquez, 299
Karla Trujillo, 603
Katelyn Teav, 644
Kathryn Leech, 749

 © ISLS

Keisuke Koba, 803
Kenneth Fei, 809
Kenneth R. Koedinger, 349
Kenneth Silseth, 589
Kevin Martin, 641
Kexin Li, 32, 695
Krista Glazewski, 354, 647
Kristina M. Lux, 309
Kristy Elizabeth Boyer, 494
Kumaran Rajaram, 23
Kylie Peppler, 758
Kyungri Kim, 629
Laila El-Hamamsy, 662
Lana Hower, 761
Lang Xu, 339
Lauren Min, 644
Lei Ding, 761
Leqi Li, 544
Leta Chan, 344
Li Wei, 70
Li-Jen Wang, 701
Liam Magee, 731
Lin Pei-Yi, 626
Lin Yung-Hsuan, 626
Linbaiyu Chen, 773
Lindsey Schulz, 414
Lingfei Cao, 776
Lingyun Huang, 524
Linor Hadar, 294
Lishan Zheng, 394
Litong Zeng, 791
Lívia Pizauro Sanchez, 812
Lixiang Gao, 722
Lu Ding, 659
Lubna Hakami, 155
Lydia Cao, 484, 577
Madhumitha Gopalakrishnan, 740
Magnus Hontvedt, 234
Manoli Pifarré Turmo, 612
Marcela Borge, 566
Marcelo Worsley, 556, 566
Margaret Ellen Seehorn, 800
Mari Longmire, 755
Maria Sitjà Martín, 612
Marianne Graves Petersen, 359
Marios Pattichis, 680
Marit Solem, 234
Mark Warschauer, 770
Marlene Scardamalia, 484, 577, 752
Martha Elena Núñez López, 419
Martin Greisel, 80, 566, 794
Maryam Etezadbrojerdi, 364
Maryann Dunn, 761
Masahiro Tada, 803
Masataka Kaneko, 665
Masaya Okada, 803
Matthew Berland, 618
Matthew Duvall, 737
Maya Israel, 494
Maya Persico, 504
Megan Humburg, 647
Megan Wiedbusch, 414, 424
Mei-Hwa Chen, 577
Melisa Orta-Martinez, 758
Melissa Özsoy, 698
Meng Wang, 224
Menglian Zhang, 710
Meral Roeben, 698
Mia Čarapina, 566
Michael Alan Chang, 556
Michael Baker, 589
Michael Tissenbaum, 165
Michael Wittmann, 549
Michelle Lui, 479
Miguel Lujan, 680
Mike Tissenbaum, 650, 791
Milouni Patel, 424
Min Li, 148
Mingyang Zhu, 710
Miroslav Suzara, 746
Mo Zhang, 148
Mon-Lin Monica Ko, 369, 556
Morten Oddvik, 234
Namrata Srivastava, 509, 647
Nanae Watanabe, 803
Natalie Rae, 566
Naveeduddin Mohammed, 509
Nga Hoang, 556, 689
Nia Nixon, 374
Noah L. Schroeder, 767
Nongnuch Ketui, 797
Nozomi Fujikawa, 464
Ole Sejer Iversen, 359
Olga Chernikova, 722
Olivia Robinson, 758
Patrick Kyllonen, 379, 668
Pattie Maes, 175
Pau Benazet i Montobbio, 707
Paula Daniela Zapata, 812
Paulo Blikstein, 743
Pedro Salinas, 304
Peter Reimann, 589
Peter W. Foltz, 556
Pierre Dillenbourg, 444
Pragati Maheshwary, 274
Qiannan Niu, 524
Qianru Lyu, 394
Quentin Biddy, 556
Rajeev Joshi, 364
Ram Das Rai, 404
Ramkumar Rajendran, 109
Rebecca G. Wells, 674
Renato Russo, 743
Richard Lee Davis, 444

 © ISLS

Richard Reeve, 484
Ritsuko Oshima, 638, 677
Robert Huang, 319, 484
Rodrigo Hidalgo, 449
Roger Azevedo, 414, 424
Rohit Raju, 556
Roland Tormey, 444
Rolf Steier, 234, 589
Rosa Mykyta-Chomsky, 119
Roshan Muddaluru, 556
Rupert Wegerif, 304, 641
Ryohei Egusa, 788
Ryunosuke Nishimura, 214
Sadhana Puntambekar, 165, 434, 653
Safinah Ali, 713
Sahana Murthy, 404
Sandra Yu Okita, 779
Sanii Mizuno, 464
Sarah Burriss, 647
Satoshi Takahashi, 788
Sebastian Simon, 484, 566
Sebastian Tews, 698
Seiyon Lee, 244, 489, 767
Seng Chee Tan, 484, 577
Serkan Ucan, 384
Seung B. Lee, 309
Shamya Karumbaiah, 274
Shan Li, 624
Shan Zhang, 244, 489, 767
Shaoming Chai, 90
Shatha N. Alkhasawneh, 99, 299
Shayan Doroudi, 3
Shi Ding, 683
Shigen Shimojo, 314, 349
Shihui Feng, 514
Shin'Ichi Konomi, 803
Shintaro Uchiyama, 464
Shiva Kalidindi, 414
Shiyan Pang, 773
Shiyu Zhang, 740
Shotaro Naganuma, 329, 334, 484, 577
Shruti Jain, 509, 647
Shuang Geng, 734
Shuang Li, 524
Shuman Wang, 746
Sidney D'Mello, 556
Sierra Rose, 556
Sophia Soomin Lee, 469
Sotheara Veng, 119
Spenser Clark, 606
Srigowri Mayasandra Prasanna, 194, 509
Stefano Cacciamani, 484
Sten Ludvigsen, 61, 589
Stéphane Allaire, 577
Stephen Elliott, 379, 668
Sudarshan Kammu, 577
Sunnie Gong, 479
Susan A. Yoon, 519, 671, 674
Sylvia Celedon-Pattichis, 680
Takeo Noda, 665
Takuya Kitamoto, 665
Tamara Sumner, 556
Tami May, 659
Tanmay Sinha, 23
Tanner Vea, 566
Teomara Rutherford, 119
Tessa Forshaw, 755
Thepchai Supnithi, 797
Thérèse Laferrière, 484, 577
Thomas Breideband, 556
Thomas Underwood, 484, 577, 761
Ti-Cheng Tsai, 701
Tianlong Zhong, 624
Tianshu Wang, 354
Tianyi Li, 129
Tianyu Ma, 51
Tongwu Jiang, 90
Tongxi Liu, 624
Toni Earle-Randell, 494
Tsubasa Minematsu, 334, 577
Ugesh Egala, 680
Umesh Timalsina, 509, 647
Uri Wilensky, 185
Venkatesh Jatla, 680
Victoria Delaney, 204, 499
Vincent Aleven, 800
Vishal Sachdev, 283
Vishwas Badhe, 109
Vitaliy Popov, 399
Vivian Leung, 577
Wakana Tsuji, 484, 752
Wangda Zhu, 773
Wei Jia, 32
Wei Yan, 719
Weiting Qi, 761
Weiyi Xiao, 635
Wen Qi, 609
Wenju Cui, 668
Wenli Chen, 3, 70, 394
William R. Penuel, 556
Wonjin Yu, 719
Xavier Ochoa, 740
Xiang Liu, 148
Xiao Zhang, 698
Xiaoge Guo, 389
Xiaojie Niu, 6
Xiaojie Zhang, 514
Xiaomeng Huang, 740
Xiaoran Li, 384
Xiaoru Li, 254
Xiaoshan Huang, 624
Xiaotian Zou, 354
Xin Guan, 544
Xingxing Xie, 656, 686

 © ISLS

Xinran Zhu, 731
Xinyue Jiao, 529
Xuanning Chen, 304, 641
Xuanyu Chen, 394
Xuechen Liu, 459
Xueqi Feng, 254, 264, 577
Xuesong Cang, 165, 434
Yang Jiang, 668
Yangbo Zhao, 264
Yanqing Sun, 577
Yasemin Erdemgil, 698
Yawen Yu, 42
Yee Yin Tan, 484, 577
Yeo Jin Kim, 647
Yeonji Jung, 469
Yi Zhang, 389
Yi-Hong Zeng, 539
Yibing Zhang, 524, 710
Yifan Jiang, 809
Yijie Lu, 389
Yimeng Sun, 524
Ying Kong, 429
Ying Xu, 459
Ying-Tien Wu, 539, 701
Yiqi Xiao, 791
Yiqin Chao, 119
Yoshihiko Kubota, 704
Yotam Hod, 484, 577
Yuanzhu Li, 224
Yubei Shi, 70
Yue Han, 609
Yue Shi, 776
Yue Zhao, 577
Yugo Hayashi, 314, 349
Yun Wen, 339
Yunseo Lee, 618
Yuqi Wang, 324
Yuqin Yang, 254, 264, 773
Yusuke Morita, 464
Yuxin Chen, 764
Yuzhe Zhou, 129
Zhen Bai, 224, 809
Zheng Bian, 674
Zhenhai He, 90
Zhenyu Cai, 444
Zhichun Liu, 776
Zifeng Liu, 529
Zongran Wang, 339

 © ISLS