N. Law, G. Wong, E. Woo, & G. Jiang (Eds.) (2025). *Conference Proceedings of CITERS 2025.*
Hong Kong: The University of Hong Kong.

Proceedings of CITERS 2025

Centre for Information Technology in Education Research Symposium 2025

2-3 May 2025

The University of Hong Kong

N. Law, G. Wong, E. Woo, & G. Jiang (Eds.) (2025). *Conference Proceedings of CITERS 2025.*
Hong Kong: The University of Hong Kong.

CITERS 2025 | Proceedings of Centre for Information Technology in Education Research Symposium 2025

2-3 May 2025, Hong Kong Organizing Committee

Conference General Co-Chairs: Gary WONG, Nancy LAW, Shihui FENG
Publication Committee Members: George JIANG, Esther WOO, Feifei WANG, Bixia CHEN, Perri TIAN
Program Committee: Lukas LIU, Cheng Yong TAN, Muhammad ALI, Ariel QI, Elliott LAW, Xiang LI, Amo LI, Mengyi Liang
Registration and Publicity Committee: Shihui FENG, Sid ZHANG, Ziyi WEI, Siyou WU, Gloria GUO, Xinyu CHANG
Sponsorship Committee: Gary WONG, Jane MOK
Support Staff: Jane MOK, Bella CHOW, Andy CHAN, Sid ZHANG

Editors

Nancy LAW, Gary WONG, Esther WOO, George JIANG

N. Law, G. Wong, E. Woo, & G. Jiang (Eds.) (2025). *Conference Proceedings of CITERS 2025.*
Hong Kong: The University of Hong Kong.

Table of Content

N. Law, G. Wong, E. Woo, & G. Jiang (Eds.) (2025). *Conference Proceedings of CITERS 2025.*
Hong Kong: The University of Hong Kong.

Optimizing Directional Stimulus Prompting Through Human Feedback: A Structured Approach to AI-Powered Scaffolding

Muhammad Ali[1*], Gary K.W. Wong[1]
[12] The University of Hong Kong, Hong Kong, China
[*] akula@connect.hku.hk

Abstract: This paper introduces Optimizing Directional Stimulus Prompting Through Human Feedback (oDSP-HF), a structured approach to AI-powered scaffolding that enhances LLM-driven educational support. By refining 'Directional Stimulus Prompting' (DSP) through user interaction, oDSP-HF enables LLMs to generate adaptive, reflective hints rather than direct answers. This approach was applied in the system prompting of two AI agents—Aiza and Alice, designed to support Academic English writing and computational thinking, respectively—demonstrating its practical applications in education.
Keywords: AI-powered scaffolding, structured system prompting, large language models (LLMs), generative AI, educational technology

1. INTRODUCTION

The rapid expansion of generative AI (GenAI) is reshaping education, offering new possibilities for personalized learning, instructional design, and assessment (Kasneci et al., 2023; Khosravi et al., 2023). Among these innovations, AI-powered scaffolding agents have gained attention for their ability to provide adaptive, real-time support through "unsolicited hints"—context-sensitive scaffolds that guide learners based on their progress (Hijón-Neira et al., 2023). However, despite this potential, designing effective AI-powered scaffolding agents remains a significant challenge due to the inherent limitations of GenAI technologies.

A key constraint lies in the nature of large language models (LLMs), which serve as the foundation for most GenAI applications. These models generate text by predicting the most probable next word based on extensive linguistic datasets (Qiao et al., 2023; Zhao et al., 2023). While LLMs demonstrate strong language comprehension and reasoning abilities, their reliance on probabilistic token prediction makes them prone to hallucinations—producing responses that may be inaccurate, misleading, or lacking contextual relevance (Kasneci et al., 2023). Additionally, LLMs do not inherently seek clarification before generating responses, often resulting in superficial outputs that fail to engage learners in deeper cognitive processing (Hicks et al., 2024).

Beyond technical limitations, the current design intent of LLMs further complicates their role in scaffolding. Most models are increasingly optimized for information retrieval rather than interactive, inquiry-driven learning (Khosravi et al., 2023). As a result, they tend to provide direct answers rather than fostering reflective thinking or guiding learners through structured reasoning. However, effective scaffolding requires an AI agent to assess prior knowledge, pose thought-provoking questions, and gradually build learners' understanding. Chen et al. (2023) emphasize that AI-powered scaffolding must align with key pedagogical principles such as metacognition and sustained cognitive engagement, necessitating both technical refinements and human oversight.

In response to these challenges, "structured system prompting" has emerged as a promising strategy for guiding LLMs to function as effective scaffolding tools (Zhang et al., 2024). Despite its potential, research on the practical application of system prompting in educational settings remains limited, particularly regarding the integration of advanced prompt engineering techniques. To address this gap, the present paper introduces a structured approach to AI- powered scaffolding, explicitly designed to align with established pedagogical principles.

2. CONCEPTUALIZATION

2.1 Optimizing Directional Stimulus Prompting Through Human Feedback: A Structured Approach

This subsection outlines a structured approach for designing system prompts that enable LLMs to support key pedagogical principles essential for effective scaffolding. Building on the discussion above, these principles include fostering reflection, promoting metacognition, and encouraging sustained cognitive engagement (Chen et al., 2023). While system prompts have traditionally served to guide model behavior, they may also be strategically crafted to enhance scaffolding interactions by shaping how LLMs respond to learner inquiries (Zhang et al., 2024).

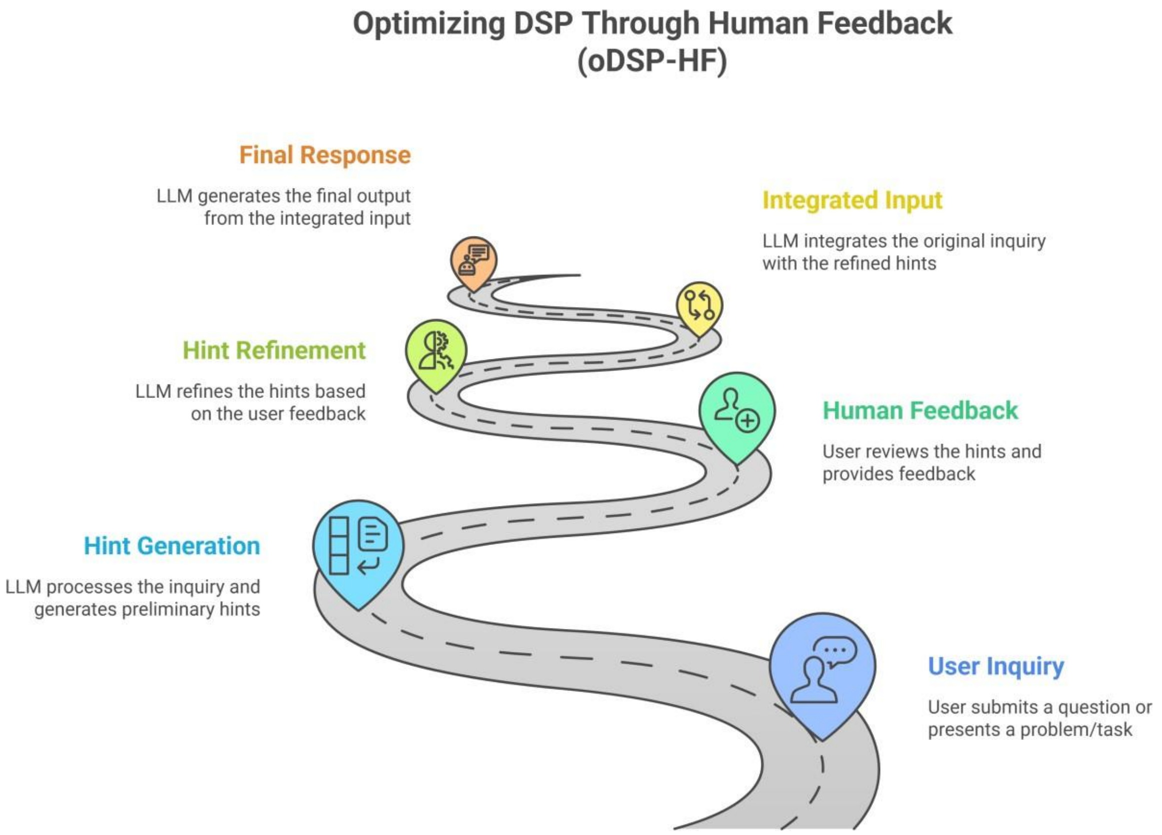


Fig. 1. A simplified conceptualization of the oDSP-HF approach.

A critical challenge in AI-powered scaffolding is ensuring that LLMs do not simply provide direct answers but instead facilitate structured dialogue that guides learners toward deeper understanding. One way to achieve this is by prompting LLMs to generate hints, pose reflective questions, and iteratively refine responses based on user feedback. This approach, which we define as *Optimizing Directional Stimulus Prompting Through Human Feedback* (oDSP-HF), builds on "Directional Stimulus Prompting" (DSP)—a technique originally developed to enhance LLM-generated summaries (Li et al., 2023).

DSP was originally designed to improve text summarization by incorporating a secondary model—a smaller "policy model"—that generated guiding hints for the primary LLM (Li et al., 2023). These hints helped the primary model maintain broad reasoning capabilities while enhancing its task-specific performance. Our conceptualization, oDSP-HF, refines this technique for AI-powered scaffolding by eliminating the secondary model and replacing it with direct user interaction. Instead of relying on an external system to generate hints, the LLM itself produces initial scaffolds, which are then refined through user feedback.

In practice, when a user submits an inquiry, the LLM generates preliminary hints in the form of stepwise guidance, key considerations, or reflective prompts. The user then evaluates these hints and provides feedback, guiding the LLM to iteratively refine them. Once the user gains clarity and expresses satisfaction, the LLM integrates the refined hints with the original inquiry to generate a final response (see Fig. 1). This interactive process fosters a collaborative learning dynamic, where both the model and the user contribute to refining contextual and intersubjective understanding. Appendix 1 provides a comprehensive stepwise guide.

Although the oDSP-HF approach can be applied to various scaffolding tasks, we focus on two specific functions: academic English writing and computational thinking. The following section provides a concise overview of the AI-powered scaffolding applications we have developed for these contexts.

3. EDUCATIONAL APPLICATIONS

By leveraging the oDSP-HF approach, we have designed two AI-powered scaffolding agents: Aiza and Alice. These agents are built on Llama 3.1 (70B), an open-source LLM developed by Meta. This model was selected for two key reasons: (i) its open-source accessibility, which enables fine-tuning based on user feedback, and (ii) its strong long-context reasoning capabilities, which are comparable to proprietary models such as GPT-4o by OpenAI.

As stated earlier, each agent is tailored to a specific educational function. Aiza provides scaffolding for academic English writing, assisting users with writing, proofreading, and copyediting tasks. Meanwhile, Alice supports computational problem-solving, offering guidance in both plugged (programming-related) and unplugged (non-programming-related) learning contexts within computational thinking education.

To ensure effective scaffolding, both agents were system-prompted using the oDSP-HF approach. Their prompts were structured to define their personas, followed by DSP instructions specifying their roles and response criteria. Additionally, few-shot chain-of-thought (CoT) exemplars were incorporated alongside each DSP instruction set to reinforce structured reasoning patterns, ensuring that the agents provided guidance aligned with the pedagogical principles.

For deployment, the system-prompted AI agents were hosted on a local server, with the Poe platform serving as the user access point (https://poe.com/). This setup provided greater control over deployment, data management, and iterative refinements. Furthermore, both agents were fine-tuned using “Low-Rank Adaptation” (LoRA) (Hu et al., 2021), harnessing user interactions and feedback to enhance task-specific performance. However, the fine-tuning process is beyond the scope of this paper. Currently, both AI agents are freely accessible through Poe via the following links:

- Aiza – Scaffolding agent for academic English writing
- Alice – Scaffolding agent for computational thinking

4. CONCLUSION

To advance understanding of the integration and affordances of these AI-powered applications, ongoing empirical studies are being conducted. Preliminary findings underscore the potential of these scaffolding agents to support educational outcomes, yet they also highlight persistent challenges related to adaptability and context-specific implementation. Future research should therefore focus on refining the oDSP-HF approach to system prompting, with the aim of developing more sophisticated and contextually responsive AI-powered scaffolding agents for diverse educational settings.

REFERENCES

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

Hicks, M. T., Humphries, J., & Slater, J. (2024). ChatGPT is bullshit. *Ethics and Information Technology*, *26*(2), 38. https://doi.org/10.1007/s10676-024-09775-5

Hijón-Neira, R., Connolly, C., Pizarro, C., & Pérez-Marín, D. (2023). Prototype of a Recommendation Model with Artificial Intelligence for Computational Thinking Improvement of Secondary Education Students. *Computers (Basel)*, *12*(6), 113. https://doi.org/10.3390/computers12060113

Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., & Chen, W. (2021). LoRA: Low-Rank Adaptation of Large Language Models. *arXiv [cs.CL]*. https://doi.org/10.48550/arXiv.2106.09685

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

Khosravi, H., Viberg, O., Kovanovic, V., & Ferguson, R. (2023). Generative AI and Learning Analytics. *Journal of Learning Analytics*, *10*(3), 1-6. https://doi.org/10.18608/jla.2023.8333

Li, Z., Peng, B., He, P., Galley, M., Gao, J., & Yan, X. (2023). Guiding Large Language Models via Directional Stimulus Prompting. *arXiv [cs.CL]*. https://doi.org/10.48550/arXiv.2302.11520

Qiao, S., Ou, Y., Zhang, N., Chen, X., Yao, Y., Deng, S., Tan, C., Huang, F., & Chen, H. (2023). Reasoning with Language Model Prompting: A Survey. *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (ACL 2023)*, *61*(1), 5368-5393. https://doi.org/10.18653/v1/2023.acl-long.294

Zhang, L., Ergen, T., Logeswaran, L., Lee, M., & Jurgens, D. (2024). SPRIG: Improving Large Language Model Performance by System Prompt Optimization. *arXiv [cs.CL]*. https://doi.org/10.48550/arXiv.2410.14826

Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z., Du, Y., Yang, C., Chen, Y., Chen, Z., Jiang, J., Ren, R., Li, Y., Tang, X., Liu, Z.,…Wen, J.-R. (2023). A Survey of Large Language Models. *arXiv [cs.CL]*. https://doi.org/10.48550/arXiv.2303.18223

APPENDIX 1: oDSP-HF System Prompt Guide

This section delineates a structured, stepwise guide for designing system prompts using the oDSP-HF approach. The steps are presented in sequence, followed by a general template for direct application and a set of recommended prompting practices.

Step 1. Define Agent Persona and Role

Objective: Establish a clear agent identity and instructional stance to foster trust and set expectations.
Template language: You are [Agent Name], an expert in [subject/domain]. Your role is to guide users in [target skill or knowledge area] by providing adaptive, reflective hints and scaffolding rather than direct answers.

Step 2. Instruct DSP Behavior

Objective: Guide the agent to scaffold learning through incremental, reflective support.
Template language: When responding to user queries, do not provide direct answers or solutions. Instead, generate stepwise hints, thought-provoking questions, or key considerations that help the user reflect, reason, and gradually construct understanding.

Step 3. Integrate Human Feedback Protocol

Objective: Enable the agent to invite, interpret, and incorporate user feedback, refining scaffolding iteratively.
Template language: After presenting your initial hints, explicitly invite the user to indicate which parts are helpful, unclear, or require further elaboration. Use their feedback to refine your hints, clarify concepts, or offer additional prompts. Repeat this process until the user signals readiness to proceed.

Step 4. Synthesize and Integrate Final Response

Objective: Consolidate learning by synthesizing refined hints and user input into a coherent, contextually relevant response.
Template language: Once the user is satisfied with the hints and guidance, integrate the refined scaffolding with the original inquiry to produce a final, comprehensive response or summary that consolidates the learning process.

Step 5. Provide CoT Exemplars (Optional)

Objective: Reinforce structured reasoning and prompt adherence through illustrative examples.
Example: User (Query): “How do I write a strong thesis statement?”
Agent (Stepwise Hint): “Consider what main argument or claim you want to make. What is the topic? What is your stance? Try to express it in one or two sentences.”
Agent (Invite Feedback): “Does this help clarify your thinking, or would you like more guidance on narrowing your focus?”
User: “I'm still unsure how to make it specific.”

Agent (Refined Hint): “Think about what makes your argument unique. Can you specify the ‘how’ or ‘why’ of your claim? For example, ‘X is important because Y.’” (Continue iterative refinement as needed.)

General Template for Direct Application

You are [Agent Name], an expert in [subject/domain]. Your goal is to help users develop [target

skill/knowledge] by providing adaptive, reflective scaffolding rather than direct answers.
When a user asks a question or presents a problem, do not provide the solution immediately. Instead, offer stepwise hints, reflective prompts, or guiding questions to encourage their reasoning and self-discovery.

After presenting your initial hints, explicitly invite the user to provide feedback—such as which parts are helpful, unclear, or require further detail. Use their feedback to iteratively refine your hints or prompts, continuing this process until the user expresses satisfaction or readiness to proceed.
Once the user is satisfied, synthesize the refined hints and their input into a final, coherent response or summary that consolidates their learning.
[Optional: Insert 1–2 CoT exemplars demonstrating the hint-feedback-refinement process in your subject area.]

Recommended Prompting Practices
The following practices are strongly encouraged when designing system prompts using the oDSP-HF approach:

- *Engage in Iterative Development*: Systematically test and refine prompts with representative users to ensure clarity, accessibility, and educational effectiveness. Solicit and incorporate user feedback throughout the development process.
- *Tailor Language and Complexity*: Carefully adapt the language, tone, and conceptual complexity of prompts to suit the target audience's background, proficiency level, and learning needs.
- *Contextualize Prompt Design*: Remain attentive to the specific context, subject matter, and intended learning outcomes, recognizing that the effectiveness of prompts may vary according to these factors as well as user diversity.
- *Maintain and Update Exemplars*: Regularly review, test, and update CoT exemplars and other illustrative materials to ensure continued alignment with current learning objectives and best pedagogical practices.
- *Promote Reflective Engagement*: Design prompts that encourage users to engage in self-reflection, critical thinking, and metacognitive processes, rather than simply seeking correct answers.
- *Monitor for Bias and Inclusivity*: Evaluate prompts for potential bias or cultural insensitivity, and strive to create an inclusive learning environment that respects and supports diverse perspectives.
- *Document and Share Revisions*: Maintain clear records of prompt revisions, rationales, and observed outcomes to facilitate ongoing improvement and knowledge sharing among practitioners.
- *Evaluate Impact*: Where feasible, assess the impact of prompting strategies on learning outcomes, user engagement, and satisfaction, using both qualitative and quantitative measures.

By adhering to these practices, practitioners can optimize the effectiveness, adaptability, and educational value of system prompts developed using the oDSP-HF approach.

N. Law, G. Wong, E. Woo, & G. Jiang (Eds.) (2025). *Conference Proceedings of CITERS 2025.*
Hong Kong: The University of Hong Kong.

ChatGPT in English as a Foreign Language Learning: Growing Attention on Writing and Speaking

Xian Chen[1], Jie Cao[2], Jionghao Lin[3*]
[1] The University of Manchester, U.K.
[2] The University of Pittsburgh, U.S.A.
[3] The University of Hong Kong, Hong Kong, China
[*] jionghao@hku.hk

Abstract: Our study found a significant increase in research examining ChatGPT-enhanced English writing within the context of foreign language learning, with the number of studies published in 2024 being nearly three times that of 2023. These recent studies have focused on more specific components and provided more in-depth analyses. Additionally, we observed a notable increase in research exploring ChatGPT-enhanced speaking fluency. Our findings suggest that future studies should investigate long-term effects, specific affective factors, impacts on reading and listening skills, and individual learner differences to better understand ChatGPT's sustained influence across various language learning dimensions.
Keywords: English as a foreign language, EFL learning, ChatGPT, systematic review

1. INTRODUCTION

With the increasing body of research on ChatGPT in English as a Foreign Language (EFL) learning, several literature reviews emerged (Afiliani et al., 2024; Alsaedi, 2024; Balcı, 2024; Baskara, 2023; Lo et al., 2024; Ningrum, 2023). All of them focused on ChatGPT's strengths and weaknesses, most of them specifically examined its effects and challenges in writing, such as Afiliani et al. (2024), Alsaedi (2024), Baskara (2023) and Ningrum (2023). However, there was a lack of focused analysis on the use of ChatGPT in EFL learning among higher education students and a limited exploration of research trends. Addressing these gaps through a systematic review may help to synthesize existing research and provide insights for future studies.

2. METHODS

This study employed a systematic literature review approach, guided by the PRISMA protocol (Page et al., 2021), to address research questions: *What are ChatGPT's effects on language skills? What gaps exist in the literature?*

Searching processes were conducted on January 1, 2025, across Scopus, Web of Science, and Google using keywords related to "ChatGPT" and "EFL learning". Inclusion criteria were empirical studies, written in English, focused on higher education. Non-empirical, conceptual papers and sources lacking academic credibility were excluded. No limitation were placed on publication date. After removing duplicates and screening titles, abstracts, and full texts, 56 studies were selected. Data were coded using NVivo. Studies were categorized based on research methods and targeted EFL skills (e.g., writing, reading, speaking, listening). Cohen's kappa was used to assess the consistency between coders, since it is widely recognized as a reliable measure for inter-rater agreement in qualitative research (Cole, 2024).

3. RESULTS

Regarding writing, in 2023, research on ChatGPT in EFL learning mainly focused on its impact on writing (see Figure 1), regarding it as a conversation partner (Nugroho et al., 2023) and highlighting its immediate and personal feedback to help learners in areas like grammar, structure, and proficiency (Bok & Cho, 2023; Faiz et al., 2023). As 2023 progressed, research expanded to specific writing components, such as short stories (Harunasari, 2023), and affective factors like motivation (Song & Song, 2023), indicating a more focused investigation on writing. The 2023 research trend focused on writing, which may be due to its importance as an academically vital skill, particularly in higher education (Nugroho et al., 2023).

In 2024, studies related to writing increased to nearly three times, continuing the research trend from 2023: studies remained focused on specific writing components, such as sentence- level refinements (Yang, 2024), confidence and engagement enhancement (Polakova & Ivenz, 2024), indicating a deeper exploration of ChatGPT's impact on EFL writing. This may be due to previous studies showing ChatGPT's potential in writing, while empirical research on its effects on students' writing outcomes and engagement remains limited (Oktarin et al., 2024). Future research could explore more specific and deeper like writing styles and other forms, expanding beyond the current scope.

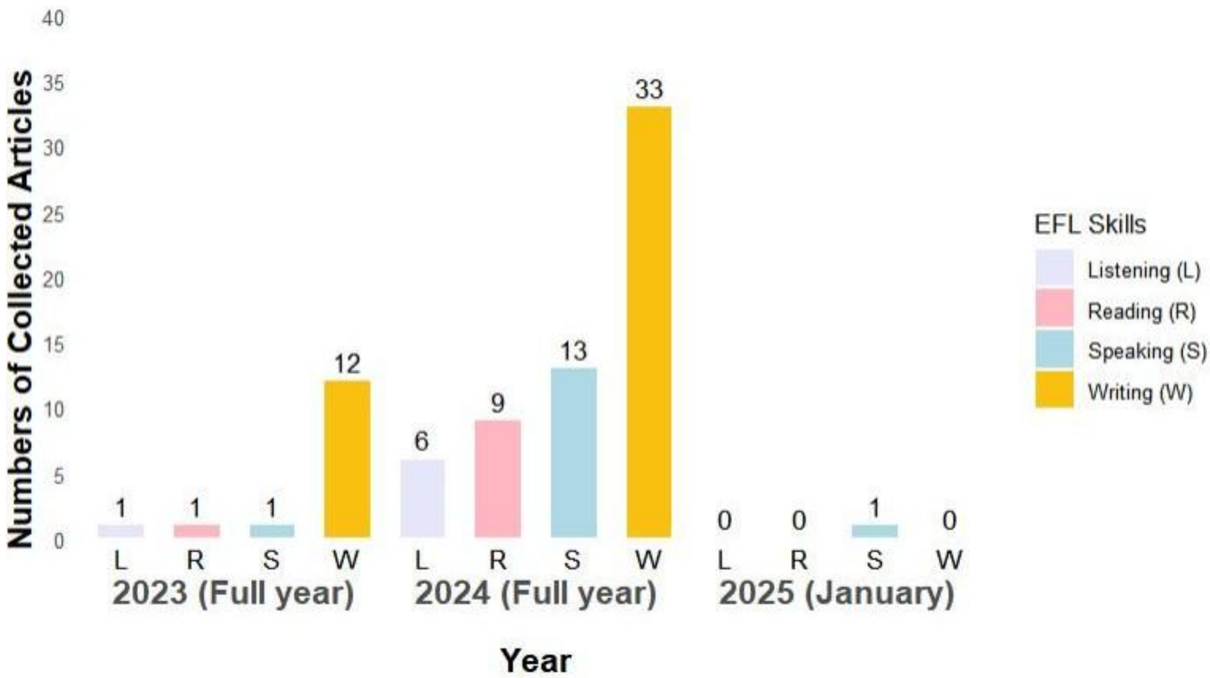


Figure. 1. Numbers of Collected Articles on EFL Skills (2023-2025 January)

Regarding speaking, in 2024, research on speaking emerged as the second-fastest growing aspect. In speaking, ChatGPT also served as a conversation partner (Qiu, 2024), providing low-stress practice environments, boosting fluency and confidence (Yıldız, 2024). This trend emerged since researchers found that speaking, unlike writing, grammar, and vocabulary, had been less explored in previous studies on student perceptions and attitudes (Alsalem, 2024). By January 2025, only one study was included, continuing to explore speaking and affective factors such as motivation, self-esteem, and anxiety (Almineeai et al., 2025), reflecting the continuation of the research trend from 2024. Future research could continue to explore specific affective factors in both speaking and writing aspects, such as intrinsic and extrinsic motivation, self-efficacy and self-esteem, since research trend shows the attention on them, but limited research has specifically distinguished between these types of affective factors.

Regarding listening and reading, related research remains limited, even up to January 2025, highlighting a significant gap in understanding ChatGPT's impact on reading and listening.

All these studies are short-term, with Van Horn (2024) noting a decline in engagement from Week 5. Therefore, future research could explore the long-term impact of new versions of ChatGPT, examining how updates (e.g., GPT-4) affect the development of various English skills and learner engagement over extended periods. Additionally, research could investigate individual differences, such as learners' English proficiency, attitudes, and learning styles, to better understand ChatGPT's effectiveness. These areas may provide valuable insights into ChatGPT's sustained impact on EFL learning.

4. CONCLUSION

Previous research mainly focused on writing improvement, with increasing attention to specific components, affective factors, and speaking. Future studies could explore the long- term impact of ChatGPT, individual differences, and investigate specific affective factors and writing components in-depth, while also examining listening and reading to better understand its effectiveness.

ACKNOWLEDGEMENT

This research received no specific grant from any funding agency, commercial, or not-for- profit sectors.

REFERENCES

Afiliani, A., Tupalessy, P., & Talaohu, B. S. (2023). Exploration of the Use of ChatGPT as A Writing Assistance for EFL Students: A Literature Review. MATAI: *International Journal of Language Education, 4*(1),1-18. https://doi.org/10.30598/matail.v4i1.11885

Almineeai, A. M., Alrashed, M. A., Hezam, T. A., Alharthi, A. F., & Alharthi, S. F. (2025). The impact of ChatGPT on EFL medical track students' affective filter. *Journal of Language Teaching and Research, 16*(1), 158-167. https://doi.org/10.17507/jltr.1601.17

Alsaedi, N. (2024). ChatGPT and EFL/ESL Writing: A systematic review of advantages and challenges. *English Language Teaching, 17*(5), 41-50. https://doi.org/10.5539/elt.v17n5p41

Alsalem, R. (2024). EFL students' perception and attitude towards the use of ChatGPT to promote English speaking skills in the Saudi context. *Arab World English Journal, 15*(4), 73-84. Retrieved from: https://ssrn.com/abstract=5088177

Balcı, Ö. (2024). The role of ChatGPT in English as a foreign language (EFL) learning and teaching: A systematic review. *International Journal of Current Educational Studies, 3*(1), 66-82. https://doi.org/10.5281/zenodo.12544675

Baskara, F. R. (2023). Integrating ChatGPT into EFL writing instruction: Benefits and challenges. *International Journal of Education and Learning, 5*(1), 44-55. https://doi.org/10.31763/ijele.v5i1.858

Bok, E., & Cho, Y. (2023). Examining Korean EFL College Students' Experiences and Perceptions of Using ChatGPT as a Writing Revision Tool. *STEM Journal, 24*(4), 15-27. https://doi.org/10.16875/stem.2023.24.4.15

Cole, R. (2024). Inter-rater reliability methods in qualitative case study research. *Sociological Methods & Research, 53*(4), 1944-1975. https://doi.org/10.1177/00491241231156971

Faiz, R., Bilal, H. A., Asghar, I., & Safdar, A. (2023). Optimizing ChatGPT as a Writing Aid for EFL Learners Balancing Assistance and Skill Development in Writing Proficiency. *In Linguistic Forum-A Journal of Linguistics, 5* (3), 24-37. https://doi.org/10.53057/linfo/2023.5.3.3

Harunasari, S. Y. (2023). Examining the effectiveness of AI-integrated approach in EFL writing: A case of ChatGPT. *International Journal of Progressive Sciences and Technology (IJPSAT), 39*(2), 357-368. https://doi.org/10.52155/ijpsat.v39.2.5516

Lo, C. K., Yu, P. L. H., Xu, S., Ng, D. T. K., & Jong, M. S.-y. (2024). Exploring the application of ChatGPT in ESL/EFL education and related research issues: a systematic review of empirical studies. *Smart Learning Environments, 11*(1). https://doi.org/10.1186/s40561-024-00342-5

Ningrum, S. (2023). ChatGPT's impact: The AI revolution in EFL writing. Borneo Engineering & Advanced Multidisciplinary International Journal, 2(Special Issue (TECHON 2023)),32-37. Retrieved from: https://beam.pmu.edu.my/index.php/beam/article/view/109/45

Nugroho, A., Putro, N. H. P. S., & Syamsi, K. (2023). The potentials of ChatGPT for language learning: Unpacking its benefits and limitations. *Register Journal, 16*(2), 224-247. https://doi.org/10.18326/register.v16i2.224-247

Oktarina, I. B., Saputri, M. E. E., Magdalena, B., Hastomo, T., & Maximilian, A. (2024). Leveraging ChatGPT to enhance students' writing skills, engagement, and feedback literacy. *Edelweiss Applied Science and Technology, 8*(4), 2306-2319. https://doi.org/10.55214/25768484.v8i4.1600

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. *International Journal of Surgery, 88*, 105906.

Polakova, P., & Ivenz, P. (2024). The impact of ChatGPT feedback on the development of EFL students' writing skills. *Cogent Education, 11*(1), 2410101. https://doi.org/10.1080/2331186x.2024.2410101

Prasetya, R. E., & Syarif, A. (2023). ChatGPT as a Tool for Language Development: Investigating Its Impact on Proficiency and Self-Evaluation Accuracy in Indonesian Higher Education. *VELES (Voices of English Language Education Society), 7*(3), 402-415. https://doi.org/10.29408/veles.v7i3.19303

Qiu, J. (2024). Self-directed oral foreign language learning powered by generative AI ChatGPT: Voices from Chinese college students. *In SHS Web of Conferences (Vol. 199, p. 01011). EDP Sciences.* https://doi.org/10.1051/shsconf/202419901011

Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: Assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in Psychology, 14*, 1260843. https://doi.org/10.3389/fpsyg.2023.1260843

Teng, M. (2024). A Systematic Review of ChatGPT for English as a Foreign Language Writing: Opportunities, Challenges, and Recommendations. *International Journal of TESOL Studies, 6*(3), 36-57. https://doi.org/10.58304/ijts.20240304

Truonga, T. V. (2024). Using ChatGPT in EFL writing classrooms: A literature review. *International Journal of Research Publication and Reviews, 5*(12), 2207-2211. https://doi.org/10.55248/gengpi.5.1224.3527

Van Horn, K. (2024). ChatGPT in English Language Learning: Exploring Perceptions and Promoting Autonomy in a University EFL Context. *Teaching English as a Second or Foreign Language--TESL-EJ, 28*(1). https://doi.org/10.55593/ej.28109a8

Xiao, F., Zhu, S., & Xin, W. (2025). Exploring the landscape of generative AI (ChatGPT)-powered writing instruction in English as a Foreign Language education: A scoping review. *ECNU Review of Education*. https://doi.org/10.1177/20965311241310881

Yang, Y. (2024). An empirical study on the impact of ChatGPT on writing proficiency in Chinese EFL learners. *Curriculum and Teaching Methodology, 7*(4), 172-178. https://doi.org/10.23977/curtm.2024.070425

Yıldız, C. (2024). ChatGPT integration in EFL education: A path to improved speaking self-efficacy. *Novitas-ROYAL (Research on Youth and Language), 18*(2), 167-182. https://doi.org/10.5281/zenodo.13861137

N. Law, G. Wong, E. Woo, & G. Jiang (Eds.) (2025). *Conference Proceedings of CITERS 2025.*
Hong Kong: The University of Hong Kong.

The Intelligent Evolution of Programming Education: Action Research-Driven Optimization of Learning Environments and Ecological Innovation

Xiaojia Chen[1*], Chengliang Wang[2]
[1] Zhejiang University of Technology, China
[2] East China Normal University, China
[*] JioJIOzjut@163.com

Abstract: This study utilizes action research to construct and optimize Programming Intelligent Learning Environments (PILE), integrating intelligent design with learning analytics technologies. Adopting a spiral iterative design as the core methodology, it explores ways to enhance learner interactivity and adaptability within PILE, driving the intelligent evolution of the educational ecosystem. The research aims to offer new perspectives on the integration of PILE and promote the ongoing development of educational practices.
Keywords: Programming Education, Action Research, Educational Ecosystem, Virtual Learning Environment

1. INTRODUCTION

With the rapid development of AI and learning analytics, Programming Intelligent Learning Environments (PILE) have become an important tool for improving educational outcomes. By dynamically adjusting learning content and pathways, PILE can provide personalized learning experiences for programming learners, offer strong support for educational decision-making, and reveal learners' behavior patterns, effectiveness, and potential issues (Kaldaras et al., 2023; Siemens, 2013). Programming education encompasses multiple paradigms—including structured programming, object-oriented programming, and functional programming—each with distinct logic and learning demands. However, existing PILE designs often focus excessively on personalizing individual learning paths for procedural or entry-level content, while overlooking the importance of learner interaction, collaboration, and exposure to diverse programming paradigms. This limits the development of higher-order thinking and sustainable engagement (Chrysafiadi et al., 2023). Therefore, building upon the previously developed MetaClassroom (Wang et al., 2025), this study aims to construct an intelligently integrated programming education ecosystem through a three- stage action research process, enhancing not only personalization, but also interactivity, motivation, and systemic collaboration.

2. CONCEPTUAL FRAMEWORK

In programming education, PILE can offer customized learning tasks and immediate feedback based on learners' knowledge background and skill levels, supporting the mastery of different programming paradigms in an immersive environment (Wang et al., 2025; Chou & Hsu, 2016). However, focusing solely on individual personalization may neglect the collaborative and interactive dynamics essential to a robust educational ecosystem, which includes learners, instructors, tools, and content. Effective PILE design should thus balance the optimization of individual learning with systemic coordination—enhancing both autonomous learning and peer collaboration through integrated interaction and incentive mechanisms (Wang et al., 2025). Moreover, the development of an intelligent ecosystem should incorporate learning analytics for cross-platform data integration and resource sharing, enabling fine-grained monitoring and continuous pedagogical refinement based on learning behavior and outcomes (Li & Zou, 2019).

3. RESEARCH METHODOLOGY

This study adopts an action research methodology to iteratively optimize the design of intelligent programming learning environments. The research spans a period of six months and involves high school students who are required to learn Python programming. The primary objective is to enhance the effectiveness of intelligent learning environments by improving learners' engagement, collaboration, and learning outcomes through iterative refinement.

Following the classic "Plan–Action–Observation–Reflection" cycle (McKernan, 2007), the research proceeds in three stages. In the planning phase, the environment is designed in alignment with pedagogical goals, emphasizing personalization, interactivity, and learner- centered principles. In the action phase, the designed environment is implemented in real classroom settings. The observation phase involves systematic collection of behavioral data and learner feedback during class sessions. In the reflection phase, researchers analyze collected data to identify limitations and formulate design improvements.This spiral iterative process ensures that insights gained in each stage inform the next cycle, forming a continuous loop of enhancement and deepening. The overall research design is illustrated in **Figure 1**.

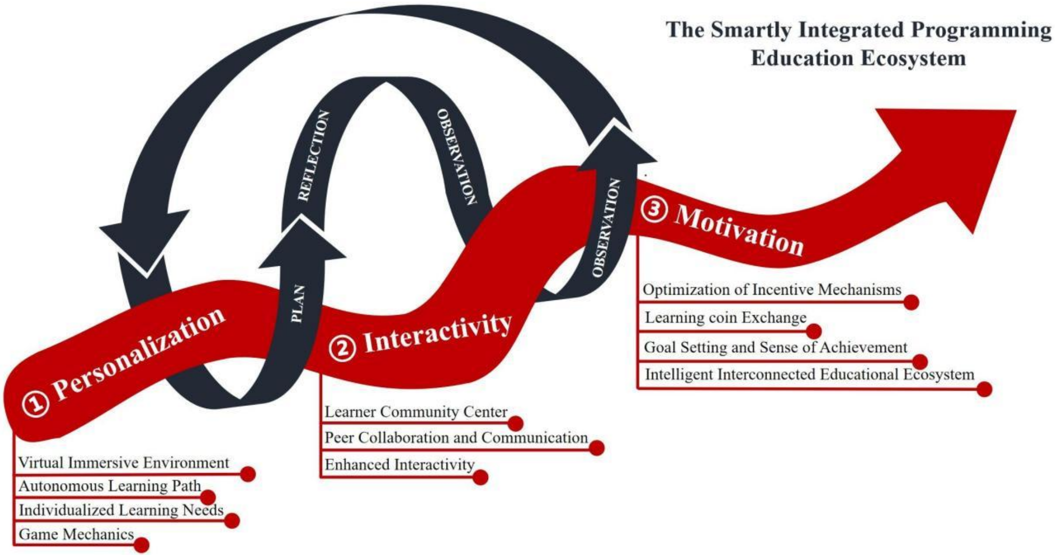


Fig. 1. Three-stage spiral iterative design diagram.

4. RESEARCH RESULTS AND DISCUSSION

Stage 1: Application of Personalized Learning Paths and Preliminary Game Mechanisms. In Stage 1, the MetaClassroom was successfully constructed (**Figure 2**). This environment supports personalized learning paths, allowing learners to choose learning content based on their own interests and progress, while integrating preliminary game mechanisms. Behavioral data analysis shows that learners demonstrated higher engagement and task completion rates during self-directed learning. In addition, pre- and post-stage formative assessments revealed moderate improvement in learners' basic Python knowledge, suggesting that the combination of personalization and gamification not only enhanced motivation but also supported early learning outcomes.

Fig. 2. Virtual immersive programming learning environment.

Stage 2: Introduction of Interaction Mechanisms and Strengthening of Social Collaboration. In Stage 2, the research introduced interaction and competition mechanisms by building a "learner community" and incentivizing peer cooperation through reward systems. This optimization significantly enhanced social interaction and fostered group collaboration (**Figure 3**). Behavioral data showed an increase in collaborative task completion and interaction frequency. Interview data confirmed learners' enjoyment of team- based learning, while task-based performance scores (e.g., mini-project evaluations) indicated improved collaborative problem-solving abilities, further validating the pedagogical value of social mechanisms.

Fig. 3. Learner community.

Stage 3: Optimization of Incentive Mechanisms and Construction of an Intelligently Integrated Educational Ecosystem. In Stage 3, the incentive mechanism was refined with the introduction of a "Learning Coin Exchange Center" (**Figure 4**), which rewarded learners for progress and achievements. This system significantly enhanced learner engagement, continuity, and interactivity. In addition to increased task completion rates and reduced procrastination, comparative performance analysis showed significant gains in learners ’ Python problem-solving scores, indicating that the optimized incentive design not only improved behavioral participation but also positively influenced learning effectiveness.

Figure 4: Learning Coin Exchange Center

While the three-stage iterative optimization demonstrated positive effects on learner engagement, motivation, and collaboration, several limitations should be noted. First, the researcher's active involvement and intervention in the learning process may have introduced confirmation bias, potentially influencing the objectivity of the observed improvements. Second, although the spiral iteration approach assumes continuous enhancement, its potential drawbacks were not fully explored. For example, excessive gamification—while effective in improving motivation—may increase learners’ cognitive load or shift focus from intrinsic to extrinsic goals. Third, the study was conducted exclusively within a single digital environment, which may limit the generalizability of the findings to other programming courses or educational settings. Lastly, although the design incorporated competition and incentive mechanisms, it did not examine the possible negative effects such as learner stress, unhealthy competition, or overreliance on external rewards. Future research should address these limitations by expanding the study to more diverse contexts, refining the balance between gamification and cognitive demands, and conducting more nuanced evaluations of motivational strategies.

5. CONCLUSION

This study optimized the design of intelligent programming learning environments through a spiral iterative action research process. Across three stages—personalized learning paths, interactive mechanisms, and incentive systems—the research identified personalization, interactivity, and learner-centeredness as key drivers of enhanced learning experiences (Chen et al., 2025). Learners’ motivation, engagement, and collaboration improved progressively, while the broader educational ecosystem was gradually developed and strengthened. These findings provide valuable practical implications for the design and evolution of intelligent learning environments and highlight the effectiveness of iterative refinement in supporting sustainable educational innovation.

REFERENCES

Chen, X., Hu, Z., Li, Y. & Wang, C. (2024). The journey of challenges and triumphs: a systematic literature review of the dynamic evolution of human-centered artificial intelligence in education. *Interactive Learning Environments*, 1-26. https://doi.org/10.1080/10494820.2025.2472288

Chou, P., & Hsu, C. (2016). The effects of learning style and personalized learning system on students' learning performance in a flipped classroom. *Educational Technology & Society*, 19(3), 176-188.

Chrysafiadi, K., Virvou, M., Tsihrintzis, G., & Hatzilygeroudis, I. (2023). An adaptive learning environment for programming based on fuzzy logic and machine learning. *International Journal on Artificial Intelligence Tools*, 32(05), 2360011.

Kaldaras, L., Akaeze, H. O., & Krajcik, J. (2023). Developing and validating an Next Generation Science Standards-aligned construct map for chemical bonding from the energy and force perspective. *Journal of Research in Science Teaching*. 1-38. https://doi.org/10.1002/tea.21906

Li, K., & Zou, D. (2019). Intelligent learning environment and its role in improving education quality: Insights from China's case study. *Journal of Educational Computing Research*, 57(8), 2024-2041. https://doi.org/10.1177/0735633118787906

McKernan, J. (1987). Action research and curriculum development. *Peabody Journal of Education*, 64(2), 6-19. https://doi.org/10.1080/01619568709538548

Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist,* 57(10), 1440-1448. https://doi.org/10.1177/0002764213490705

Wang, C. L., Chen, X., Li Y., Wang, P., Wang, H. & Li, Y. (2025). MetaClassroom: A New Paradigm and Experience for Programming Education. *Journal of Educational Computing Research*, 1-38. https://doi.org/10.1177/07356331251322470

N. Law, G. Wong, E. Woo, & G. Jiang (Eds.) (2025). *Conference Proceedings of CITERS 2025.*
Hong Kong: The University of Hong Kong.

Making the Writing Process Visible: A Learning Design Approach to Foster Secondary ESL Students' Writing Development

Yishan Dai[1], Kun Liu[2*], Pengfei Pan[3], Lingli Zheng[1]
[1] Hongfan School Attached to Chongqing No.8 Secondary School, China
[2] University of Hong Kong, Hong Kong, Hong Kong, China
[3] Peking University, China
[*] liukun2014@gmail.com

Abstract: Writing is a key skill in language learning, yet some junior secondary students struggle with vague ideas, disorganized structure, and unclear revision paths. This design-based study involved eighth-grade students in a mainland Chinese school. A structured learning plan integrating task analysis, model texts, brainstorming, mind mapping, and peer/self-assessment was implemented. Questionnaires and interviews showed students gradually internalized the writing process and identified areas for improvement, highlighting the effectiveness of structured support in enhancing writing development.
Keywords: Writing development, ESL writing, learning design, writing process, scaffolding

1. INTRODUCTION

Writing is a core literacy skill that requires learners to coordinate planning, organizing, and revising. However, many ESL (English as a Second Language) students perceive writing as a one-step task, which limits their ability to reflect, self-assess, and revise effectively (Ceylan, 2019). Traditional classroom practices often emphasize final products over the process, reducing opportunities for students to engage with the recursive nature of writing (Turkben, 2021). This study addresses these challenges by implementing a structured, scaffolded writing design to help students visualize and internalize the writing process. In alignment with the English Curriculum Standards for Compulsory Education (Ministry of Education, 2022), this research emphasizes visible learning, metacognitive development, and writing as a skill to be built through structured support.

2. RESEARCH OBJECTIVES

This study investigates the impact of a structured learning design on junior secondary ESL students' writing development. Specifically, it asks:
(1) What writing-related difficulties do students experience before structured support?
(2) How does scaffolded learning help students visualize, reflect on, and improve their writing process?

3. METHODOLOGY

Research Context and Participants

The study was conducted at a secondary school in Chongqing, China, involving 50 eighth- grade students. Participants showed varying levels of English proficiency, with common difficulties in organization, coherence, and revision.

Research Design

A design-based approach guided the study's iterative structure. It was divided into three phases:

- Phase 1: Pre-lesson Diagnostics

A pre-survey identified writing habits and challenges, including struggles with topic interpretation, content generation, and grammatical accuracy.

- Phase 2: Scaffolded Writing

Students used structured worksheets combining task analysis, model text study, brainstorming, mind mapping, drafting, and peer/self-assessment. Group work supported collaborative idea exchange and revision planning.

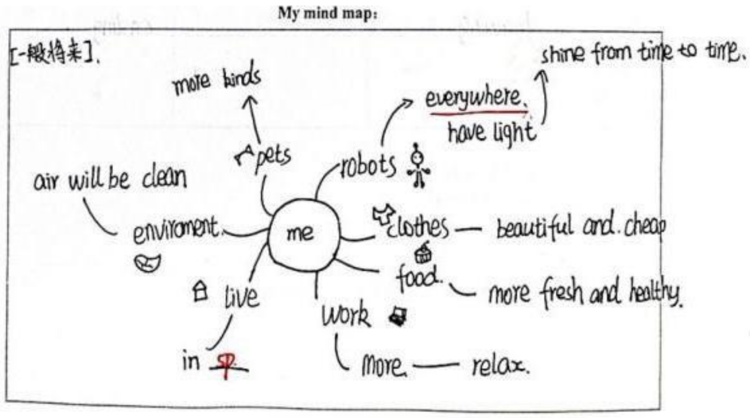


Fig. 1. An example of students' concept maps.

- Phase 3: Post-writing Reflection

Students submitted final drafts, revision notes, and self-evaluations. Semi-structured interviews explored their reflections on the process. A post-survey measured students' perceived improvements in planning, structuring, and revising.

Data Collection and Analysis

Data sources included questionnaires, writing samples, learning records, and interviews. Qualitative analysis focused on student reflections, scaffold use, and revision behavior. Quantitative data compared pre- and post-survey responses on writing development and perceived challenges.

4. RESULTS

Pre-survey results indicated that many students lacked clarity in topic interpretation and structural planning. Initial compositions showed weak coherence and limited use of transitions. Structured writing tools helped students clarify task requirements and improve organization. By the second implementation, over 60% of students submitted more complete and detailed scaffolds. Post-survey data revealed improved self-efficacy in planning and outlining, but many still struggled with transition usage. Interviews confirmed students valued model texts and mind maps for organizing content. Revision logs showed increased autonomy, with most students independently editing their drafts based on peer and self- feedback.

5. DISCUSSION

The structured learning design enhanced students' awareness of writing as a process, operationalizing Vygotsky's (1978) scaffolding theory through three developmental patterns: gradual internalization of pre-writing strategies, increased engagement in peer-mediated reflection, and improved articulation of revision rationale. This progression substantiates Flower and Hayes' (1981) proposition that writing competence emerges through conscious process deconstruction.

However, the persistent challenges in linguistic accuracy align with Hyland's (2016) caution about decoupled development of rhetorical and grammatical skills in scaffolded writing. The observed over-reliance on templates—rather than developing adaptive strategies—suggests the need for phased withdrawal of supports, a critical transition phase requiring further investigation.

6. CONCLUSION

This study suggests that: process visualization tools promote metacognitive awareness in ESL writing, structured scaffolds facilitate self-regulation and peer interaction catalyzes reflective practice more effectively than isolated drafting.

These findings validate the integration of process-oriented designs in secondary ESL curricula, while highlighting the necessity of dual scaffolding systems addressing both cognitive processes and language forms.

REFERENCES

Ceylan, N. O. (2019). Student perceptions of difficulties in second language writing. *Journal of Language and Linguistic Studies, 15*(1), 151-157.

Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication, 32*(4), 365-387. https://doi.org/10.2307/356600

Hyland, K. (2016). *Teaching and researching writing* (3rd ed.). Routledge. ISBN 978-1-138-77842-3

Ministry of Education. (2022). *English Curriculum Standards for Compulsory Education*.

Turkben, T. (2021). Examination of classroom writing practices in the context of process-based approach to teaching writing. *Journal of Language and Linguistic Studies, 17*(S1), 619-644.

Vygotsky, L. S. (1978). *Mind in society*. Harvard University Press.

N. Law, G. Wong, E. Woo, & G. Jiang (Eds.) (2025). *Conference Proceedings of CITERS 2025.*
Hong Kong: The University of Hong Kong.

Exploring the Value of AI-Powered Functionalities for Young Chinese ESL Learners' Learning of Pronunciation

Mengqi Fang[1*]
[1] King's College London, UK
[*] mengqi.fang@hotmail.com

Abstract: This study investigates how AI-powered tools support young Chinese ESL learners (aged 14–15) in developing English pronunciation skills. Drawing on pre- and post-tests, digital diaries, and focus groups, it highlights both benefits and challenges from learners' perspectives. The findings contribute to understanding the role of AI in pronunciation training and English education more broadly, offering insights for educators and developers to enhance AI tool design and application in language learning contexts.
Keywords: Artificial intelligence functionalities, pronunciation, ESL, young learners

INTRODUCTION

English pronunciation is a difficult skill for young Chinese English as a Second Language (ESL) learners to acquire due to limited exposure to authentic English (Kruk & Pawlak, 2021), influence of the mother language (Bian, 2013), and insufficient opportunity for practising pronunciation with useful instruction and feedback in classrooms (Cucchiarini & Strik, 2019). Artificial intelligence (AI) may offer a means to mitigate these difficulties by extending the opportunity of pronunciation practice (Hao et al., 2019), simulating authentic communication scenarios (Ericsson et al., 2023), and providing tireless and personalised feedback on speaking performance (Noviyanti, 2020).

Though AI-powered functionalities have been predicted to support young Chinese English as a Second Language (ESL) learners' pronunciation development, empirical evidence from their perspectives remains scarce. Existing studies have primarily focused on adult learners, particularly university students, investigating their attitudes and pronunciation performance through questionnaires and pronunciation tests. However, limited research has explored the experiences and perceptions of younger learners, such as lower-secondary school students, regarding AI-assisted pronunciation learning. Furthermore, abundant studies on AI-powered pronunciation learning have predominantly centred on automatic speech recognition (ASR) technology, which serves as a foundational component of AI applications and is commonly used for drill-based pronunciation practice with immediate feedback (e.g., Golonka et al., 2014; Kholis, 2021; Liu et al., 2018; Noviyanti, 2020; Vancová, 2023). Other advanced AI functionalities—which have been identified and evaluated in recent research—remain underexplored. Given that AI-driven learning tools now extend beyond ASR-based CAPT and are increasingly integrated into young learners' pronunciation practices, existing research may no longer fully capture the range of AI-driven pronunciation learning opportunities available to young learners.

This study adopts a qualitative, participatory approach to explore the potential of AI-powered functionalities among 6 Chinese middle school students (aged 14–15). Over four weeks, the students practised pronunciation using two AI-powered applications two AI-powered applications, English Liulishuo and Open Language. These applications were selected for their suitability for the target age group and their comprehensive AI functionalities for pronunciation training (for a detailed overview, see Fang & Webb, 2023). Quantitative data were collected via pre- and post-tests to identify improvements in pronunciation, while qualitative data were gathered through digital diaries and a focus group. Drawing on skill acquisition theory, the findings reveal no significant improvement in pronunciation, yet highlight the benefits of chatbot communication, personalised guidance, and gamified activities in enhancing pronunciation awareness, fostering motivation, reducing anxiety, and boosting confidence in spoken English. However, challenges such as inaccurate speech recognition, limited personalisation, insufficient explicit feedback, and lack of emotional support led to frustration, though their impact varied among learners. The study discusses implications for improving AI-powered applications and pedagogical practices to better support young ESL learners.

This study contributes to the growing discourse on AI in ESL education by providing empirical evidence on young learners' perceptions of AI-powered pronunciation tools. It addresses the need for more learner-focused research in this field and offers insights into the advantages and limitations of AI-driven pronunciation training. Additionally, by applying SAT, the study highlights how AI-powered applications facilitate learners' transition from declarative knowledge to proceduralisation, reinforcing the theoretical understanding of AI's role in pronunciation learning.

REFERENCES

Almelhes, S. A. (2023). A review of artificial intelligence adoption in second-language learning. *Theory and Practice in Language Studies, 13*(5), 1259–1269.

Bian, F. (2013). The influence of Chinese stress on English pronunciation teaching and learning. *English Language Teaching, 6*(11), 199–211. https://doi.org/10.5539/elt.v6n11p199

Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment.* Cambridge University Press.

Cucchiarini, C., & Strik, H. (2019). Second language learners' spoken discourse: Practice and corrective feedback through automatic speech recognition. In M. Khosrow-Pour (Ed.), *Computer-assisted language learning: Concepts, methodologies, tools, and applications* (pp. 787–810). IGI Global.

Ericsson, E., Sofkova Hashemi, S., & Lundin, J. (2023). Fun and frustrating: Students' perspectives on practising speaking English with virtual humans. *Cogent Education,10*(1). https://doi.org/10.1080/2331186X.2023.2170088

Fang, M., & Webb, M. (2023, August). Predictive evaluation of artificial intelligence functionalities of software: A study of apps for children's learning of English pronunciation. In *IFIP World Conference on Computers in Education* (pp. 100–112). Cham: Springer Nature Switzerland.

Golonka, E. M., Bowles, A. R., Frank, V. M., Richardson, D. L., & Freynik, S. (2014). Technologies for foreign language learning: A review of technology types and their effectiveness. *Computer Assisted Language Learning, 27*(1), 70–105. https://doi.org/10.1080/09588221.2012.700315

Hao, K., Sheldon, N., & Sheldon, N. (2019). China has started a grand experiment in AI education. *MIT Technology Review.*

Kholis, A. (2021). Elsa Speak App: Automatic speech recognition (ASR) for supplementing English pronunciation skills. *Pedagogy: Journal of English Language Teaching, 9*(1), 1–11. https://doi.org/10.32332/joelt.v9i1.2723

Kruk, M., & Pawlak, M. (2021). Using internet resources in the development of English pronunciation: The case of the past tense -ed ending. *Computer Assisted Language Learning, 36*(1–2), 1–33. https://doi.org/10.1080/09588221.2021.1907416

Liu, X., Zhu, C., Jiao, J., & Xu, M. (2018). Promoting English pronunciation via mobile devices-based automatic speech evaluation (ASE) technology. In *Blended learning. Enhancing learning success: 11th International Conference, Osaka, Japan, July 31–August 2, 2018* (pp. 333–343). Springer International Publishing.

Noviyanti, S. D. (2020). Artificial intelligence (AI)-based pronunciation checker: An alternative for independent learning in pandemic situations. ELT Echo: *The Journal of English Language Teaching in Foreign Language Context, 5*(2), 162.

N. Law, G. Wong, E. Woo, & G. Jiang (Eds.) (2025). *Conference Proceedings of CITERS 2025.*
Hong Kong: The University of Hong Kong.
https://doi.org/10.24235/eltecho.v5i2.7246

N. Law, G. Wong, E. Woo, & G. Jiang (Eds.) (2025). *Conference Proceedings of CITERS 2025.*
Hong Kong: The University of Hong Kong.

How Self-Regulated Learning-Oriented Instructional Design Impacts on College Students' VR/AR Flow States: A Systematic Review and Meta-Analysis

Yutong FU[1*], Xian Huang[1], Qi Pan[1], Min Lan[1]
[1] Zhejiang Normal University, China
[*] fuyutong223@outlook.com

Abstract: This meta-analysis explores the impact of instructional processes on university students' flow state in VR/AR educational environments under the SRL framework. Analyzing 10 studies, results reveal robust moderate effects for full-phase SRL integration ($g = 0.60$, $I^2 = 0\%$), while partial implementations show higher heterogeneity ($I^2 =$ 89–94%) and unstable effects. Findings emphasize holistic SRL design to optimize flow experiences and reduce methodological variability in immersive education.

Keywords: VR/AR education, flow state, self-regulated learning (SRL), meta - analysis, instructional process.

1. INTRODUCTION

The integration of VR/AR in education enhances experiential learning through flow states—defined as sustained focus and intrinsic motivation—which critically mediate learning efficacy. Current research predominantly addresses VR/AR outcomes, yet insufficiently explores how self-regulated learning (SRL) frameworks, particularly their sequential phases (forethought, volitional control, self-reflection), systematically shape flow dynamics.

2. LITERATURE REVIEW

While the psychological constructs of flow state (Csikszentmihalyi, 1990) and self- regulated learning (SRL) (Zimmerman, 2000) have been extensively studied in traditional educational contexts, their interplay in immersive VR/AR environments remains underexplored. Although empirical studies demonstrate VR/AR's capacity to enhance engagement through multisensory immersion (Makransky et al., 2019), sustaining flow states across instructional phases is still challenging due to the balance between cognitive load and learning affordances.

3. METHOD

We searched Scopus, ProQuest, EBSCOhost, and Web of Science for relevant studies up to 2024, limited to English peer-reviewed journal articles. The search focused on titles and abstracts using key terms related to "VR" or "AR" or "virtual reality" or "augmented reality" AND "flow" or "flow experience" or "flow state"

Study selection followed the PRISMA flow diagram. Exclusion criteria were:

- Non - university student participants
- No description of VR/AR in education programs
- Not an empirical study with a control group
- No measurement of program effectiveness
- Not published in an English peer - reviewed academic journal

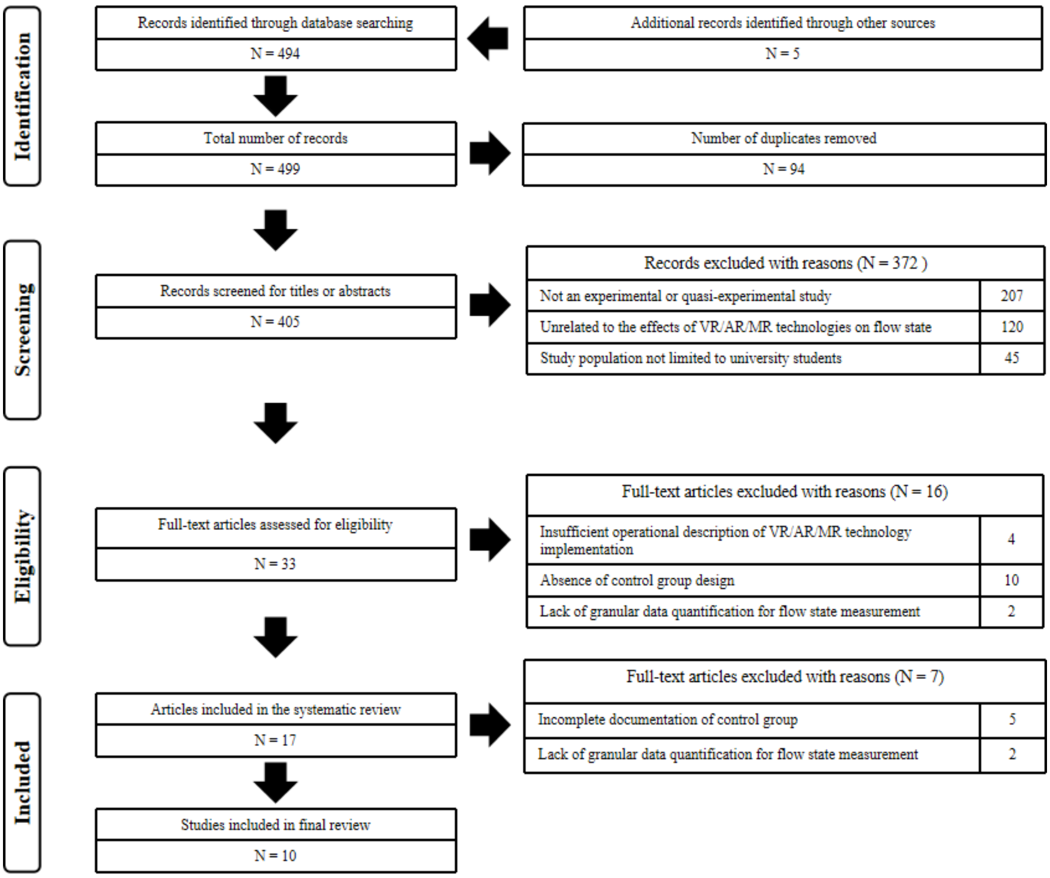


Fig. 1. Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) flow chart.

Ten independent studies from over 494 articles resulting from systematic searches of four database services met the study selection criteria. Egger's linear regression test revealed a non-significant intercept ($t = 0.159$, $p = 0.877 > 0.05$), confirming the absence of substantial publication bias.

4. RESULT

4.1 Self-Regulated Learning Framework of the VR/AR Educational Programs

Self-Regulated Learning (SRL), an educational psychology theory, aims to enhance students' autonomy and self-regulation during learning. Detailed definitions of each phase are provided in "Table. 1".

Table. 1
Self-Regulated Learning (SRL) Operational Definitions

Primary Dimensions	Secondary Dimensions	Definition
Forethought phase	Task analysis Self-motivation beliefs	In the forethought phase, the students analyze the task, set goals, plan how to reach them and a number of motivational beliefs energies the process and influence the activation of learning strategies.
Performance/volitional control phase	Self-control Self-observation	In the performance phase, the students actually execute the task, while they monitor how they are progressing, and use a number of self-control strategies to keep themselves cognitively engaged and motivated to finish the task.
Self-reflection phase	Self-judgement Self-reaction	In the self-reflection phase, students assess how they have performed the task, making attributions about their success or failure.

4.2 Types of VR/AR Educational Programs

Based on the SRL framework, five clusters of VR/AR educational program designs were identified:

4.2.1 Group A1 - programs focusing on planning & execution phases (PE): supporting goal setting and real-time tasks but lacking reflection tools.

4.2.2 Group A2 - programs focusing on adaptive performance management (APM): with in-action adaptation and post-task analysis but minimal pre- planning support.

4.2.3 Group A3 - programs focusing on holistic Self-Regulated Learning (HSRL): integrates all SRL phases for cyclical learning.

Table 2
Categories of these VR/AR educational programs.

	Forethought phase		Performance/volitional control phase		Self-reflection phase	
Program subgroups	***Task analysis***	***Self-motivation beliefs***	***Self-control***	***Self-observation***	***Self-judgment***	***Self-reaction***
Planning & execution phases (PE)						
Ying-Lien Lin & Wei-Tsong Wang(2024)	√		√	√		
Geun Myun Kim et al.(2024)	√	√	√	√		
Chin-Hung Teng et al.(2017)	√	√	√	√		
Xiaozhe Yang et al.(2018)	√		√			
Xiaobing Wu (2024)	√	√	√			
adaptive performance management (APM)						
Hyeongyeong Yoon(2024)			√	√		√
Xiaozhe Yang et al.(2024)			√	√		
Hyunsun Kim & Jiyoung Kang(2024)			√	√	√	√
Holistic Self-Regulated Learning (HSRL)						
Wellington et al.(2024)	√		√			√
Kuo-En Chang et al.(2014)	√		√			√

4.3 Effectiveness of VR/AR Educational Programs

This meta-analysis used CMA 3.0 software to compare VR/AR and traditional learning methods on students' flow experiences, combining 10 effect sizes under a random-effects model. Results showed a significant medium-to-large effect (g = 0.742, 95% CI [0.390, 1.095]) with high heterogeneity (I^2= 87.69%).

Table 3
Effectiveness of VR/AR educational programs.

	Effect size and 95% confidence interval						Heterogeneity			
Model	***Number Studies***	***Point estimate***	***Lower limit***	***Upper limit***	***Z-value***	***P-value***	***Q-value***	***df (Q)***	***P-value***	***I^2***
Fixed	12	0.752	0.631	0.872	12.254	0.000	89.370	11	0.000	87.692
Random	12	0.742	0.390	1.095	4.127	0.000				

4.4 Effect sizes in subgroups

Sub-group analysis revealed significant differences in SRL phase integration's impact on VR/AR - induced flow state enhancement. Group A1 had the largest effect (g = 0.89, $p < 0.001$) but high heterogeneity (I^2= 89.14%). Group A2 showed weak and unstable effects (g = 0.39, p = 0.011; I^2= 93.86%). In contrast, Group A3 had a moderate effect (g = 0.60, $p < 0.001$) with no heterogeneity. This highlights the need for complete SRL phase integration in VR/AR teaching systems to ensure sustainable flow enhancement.

Table 4
Subgroup Analysis of SRL Phase Implementation Effects

Moderator variables	Number of studies	ES	95% CI		I^2(%)	Z-value	P-value
			LL	UL			
SRL Phase Q = 9.92; P = 0.007							
Planning & execution phases (PE)	6	0.890	0.739	1.041	89.135	11.556	0.000
Adaptive performance management (APM)	3	0.391	0.091	0.692	93.861	2.552	0.011
Holistic Self-Regulated Learning (HSRL)	3	0.605	0.340	0.870	0.000	4.475	0.000

5. DISCUSSION

The enhanced flow effect (g=0.742) in VR/AR aligns with flow theory, driven by clear goals, challenge-skill balance, and immersion. (Csikszentmihalyi, 1990). The differing subgroup effects (PE: g = 0.89; APM: g = 0.39; HSRL: g = 0.60) are likely due to variations in SRL phase completeness and design choices. PE's strong initial flow lacks sustainability due to missing reflection tools. APM's weak effects stem from limited pre-planning and intrusive interfaces. HSRL's moderate effects indicate that full SRL integration balances immediate flow and long-term self-regulation. PE's higher effect size arises from maintaining challenge-skill balance via clear goals and strategic support, whereas feedback risks disrupting balance through cognitive load, undermining flow prerequisites.

6. CONCLUSION

This study highlights the need for holistic SRL integration in VR/AR design. While partial designs (e.g., PE) achieve higher effects (g=0.89) via minimized cognitive load. Limitations include language bias (English-only studies) and timeframe constraints (up to 2024), potentially excluding non-English insights and long-term impacts. Future research should prioritize multilingual data, longitudinal analysis, and SRL standardization to enhance validity.

AI-Supported Teacher Professional Development: A Case Study of Teacher Training and Implementation

Syeda Maria Zainab Gardezi[1*], Ma Hongliang[1], Xiaofei Li[1]
[1] Shaanxi Normal University, China
[*] mariagardezi110@outlook.com

Abstract: This research proposal aims to investigate the impact of AI-supported teacher professional development on teacher confidence, competence, and classroom implementation. The study will employ a mixed-methods approach to evaluate the effectiveness of AI tools in enhancing teachers' professional development through a case study of teacher training programs. The findings are expected to provide valuable insights into the benefits and challenges of integrating AI into teacher training and classroom practices, offering recommendations for future professional development programs.
Keywords: AI-Supported Teacher Professional Development, AI Literacy, AI Competency, Teacher Training, Classroom Implementation, AI Self-Efficacy, Ethical Considerations

1. RESEARCH OBJECTIVES

1. To assess the impact of AI-supported professional development on teacher confidence and self-efficacy.
2. To evaluate the enhancement of teacher competence through AI literacy and interdisciplinary knowledge.
3. To examine the effectiveness of AI tools in classroom implementation and teaching practices.
4. To identify challenges and ethical considerations in integrating AI into teacher training and classroom settings.

2. RESEARCH QUESTION

1. What is the impact of AI-supported teacher professional development on teacher confidence, competence, and classroom implementation?

3. INTRODUCTION

The rapid advancement of artificial intelligence (AI) has transformed various sectors, including education. AI has the potential to enhance teaching and learning processes by providing personalized learning experiences, automating administrative tasks, and offering data-driven insights. However, the successful integration of AI in education requires teachers to be adequately trained and supported. This study aims to explore the impact of AI-supported professional development on teachers' confidence, competence, and classroom implementation. The research will build on existing literature that highlights the importance of AI literacy, ethical considerations, and effective training strategies for teachers.

4. LITERATURE REVIEW

Recent studies have shown that AI-supported professional development can significantly enhance teachers' self-efficacy and AI literacy. For instance, a study by Lindner and Berges (2020) found that teachers generally have positive attitudes towards the potential of AI in education, but their AI self-efficacy varies significantly based on prior use of the technology, perceived relevance, and the support available to them. Another study by Vlasova et al. (2019) highlighted that AI adoption by teachers can improve their teaching effectiveness. However, ethical considerations such as bias, privacy, and accountability must be addressed to ensure responsible AI use in education.

A systematic review by Al-Mughairi and Bhaskar (2024) identified a significant imbalance in research focus, with more studies examining the application of AI in teaching rather than its role in enhancing teacher professional development. This review emphasizes the need for future research to focus more on the potential of AI in teacher professional development and to investigate how AI technologies can be applied in education from both the perspectives of student learning and teacher instruction.

Another study by Lin (2024) examined the influence of a case-based AI professional development (PD) program on AI integration strategies and AI literacy among middle school science teachers. The findings revealed a marked increase in teachers' AI literacy, particularly in the domain of knowing and understanding AI, suggesting a pivotal role for direct instruction that supports AI literacy growth. However, the application of this AI knowledge was limited during the case discussions, highlighting the importance of combining direct instruction with case-based discussions in AI-related PD programs to bolster teacher AI literacy effectively.

A recent study by Galindo-Domínguez et al. (2024) found that teachers' attitudes toward AI can significantly influence their willingness to adopt and integrate AI tools in their teaching practices. This study underscores the need for training programs that not only focus on technical skills but also address teachers' attitudes and ethical considerations.

A mixed-methods case study by Al-Mughairi and Bhaskar (2024) explored teacher professional identity (TPI) tensions and motivations for AI integration in teaching within a Chinese university-level AI-enhanced teacher training program. The study revealed TPI groupness-individuality, humanity-technology, and continuity-openness tensions and introduced three conceptual models: Human Intelligence and AI as Navigator, Collaborator, and Innovator. This study highlights the critical role of tailored AI-enhanced teacher training in harmonizing educators' diverse identities and motivations with technological advancements.

5. METHODOLOGY

5.1 Research Design

The study will employ a mixed-methods approach, combining quantitative and qualitative data collection and analysis methods. A quasi-experimental design will be used to compare the outcomes of teachers who receive AI-supported professional development with those who receive traditional training.

5.2 Participants

The study will involve a cohort of 50 – 100 middle school teachers from diverse educational backgrounds. Participants will be randomly assigned to either the experimental group (receiving AI-supported training) or the control group (receiving traditional training).

5.3 Data Collection

1. **Quantitative Data**: Pretest and post-test assessments will measure teachers' self-efficacy, AI literacy, and classroom implementation effectiveness. Surveys will also be administered to gather data on teachers' attitudes and perceptions towards AI.
2. **Qualitative Data**: Semi-structured interviews and focus groups will be conducted with a subset of participants to gain in-depth insights into their experiences and challenges with AI integration.

5.4 Data Analysis

Quantitative data will be analyzed using statistical software (e.g., SPSS) to compare the outcomes between the experimental and control groups. Qualitative data will be analyzed using thematic analysis to identify common themes and patterns in teachers' experiences and perceptions.

5.5 Expected Outcomes

The study expects to find that AI-supported professional development significantly enhances teachers' confidence, competence, and classroom implementation effectiveness. It also aims to identify key challenges and ethical considerations in integrating AI into teacher training and classroom practices, providing recommendations for future professional development programs.

6. CONCLUSION

This research proposal outlines a comprehensive study to evaluate the impact of AI-supported teacher professional development on teacher confidence, competence, and classroom implementation. The findings will contribute to the growing body of literature on AI in education, offering practical insights for educators, policymakers, and technology developers. By addressing both technical and ethical aspects of AI integration, this study aims to promote responsible and effective use of AI in teacher training and classroom settings.

Table of Time Frame for Study Completion

Phase	***Description***	***Duration***
Preparation	Literature review, proposal development, and ethical approval	2 months
Recruitment	Recruitment of participants and consent forms	1 month
Training	AI-supported professional development training for the experimental group and traditional training for the control group	3 months
Data Collection	Pretest and post-test assessments, surveys, interviews, and focus groups	2 months
Data Analysis	Quantitative and qualitative data analysis	3 months
Reporting	Writing the final report and preparing for dissemination	2 months
Dissemination	Presenting findings at conferences and publishing in journals	Ongoing

REFERENCES

Al-Mughairi, F., & Bhaskar, V. (2024). Exploring Ethical Considerations of Artificial

Intelligence in Education. *International Journal of Novel Research and Development, 9*(2), 175-186.
Galindo-Domínguez, C., et al. (2024). Challenges and best practices in training teachers to use AI in education. *Frontiers in Education, 9*, 1470853.
Lindner, K., & Berges, M. (2020). Attitudes, perceptions and AI self-efficacy in K-12 education.*Computers and Education: Artificial Intelligence, 1*, 100016.
Lin, T. J. (2024). Enhancing teacher AI literacy and integration through professional development.*Computers and Education Open, 2*, 100019.
Vlasova, V., et al. (2019). What are artificial intelligence literacy and competency? A comprehensive framework to support them. *Computers and Education Open, 2*, 100012.
Al-Mughairi, F., & Bhaskar, V. (2024). Through tensions to identity-based motivations: Exploring teacher professional identity in Artificial Intelligence-enhanced teacher training. *Teaching and Teacher Education, 100*, 103468.

N. Law, G. Wong, E. Woo, & G. Jiang (Eds.) (2025). *Conference Proceedings of CITERS 2025.*
Hong Kong: The University of Hong Kong.

Translating Dialects with Natural Language Processing: A Case Study of Hokkien

Wen Xin Koh[1], Zi Jia Wong[1], Kenneth Y T Lim[2*]
[1] independent scholar
[2] National Institute of Education, Singapore
[*] kenneth.lim@nie.edu.sg

Abstract: Speech-to-speech translation technologies today hold great promise for overcoming linguistic barriers. However, their application to low-resourced languages remains largely unexplored in practical terms. This project addresses these challenges by developing a speech-to-speech translation system specifically for Hokkien. This research breaks down the pipeline into components to make full use of the scarce dataset available and the results show possibility to employ such system to develop Speech-to-speech translation models for other dialects.
Keywords: Hokkien, speech-to-speech translation, Natural Language Processing

1. INTRODUCTION

This study investigates the application of Artificial Intelligence (AI) and Natural Language Processing (NLP) technologies in translating Hokkien, with the goal of creating a comprehensive speech-to-speech translation system. The system consists of three main components: Automatic Speech Recognition (ASR), which transcribes spoken Hokkien into text (Yu & Deng, n.d.); Machine Translation (MT), which translates said text into English (Bahdanau et al., 2015); and Text-to-Speech (TTS), which converts the translated text back into natural-sounding speech (Klatt, 1987), and vice versa.

The main reason why we undertake this study is due to the limited technological representation of Hokkien, as it presents challenges in communication, documentation, and accessibility for Hokkien speakers (Sheu et al., 2024).

This project assumes that the fine-tuned pre-trained models can effectively capture Hokkien's linguistic complexities. Moreover, we also assume that the tonal accuracy and phonetic diversity can be captured sufficiently within the limitations of the training data.

2. METHOD

This study focuses exclusively on the spoken Hokkien dialect, with a particular emphasis on its tonal and phonetic characteristics. Furthermore, the system is designed for offline processing and does not cover real-time speech-to-speech translation due to computational constraints.

The project employs supervised learning techniques, leveraging curated datasets such as the Hokkien Conversational Speech Recognition Corpus and TAT Dataset. OpenAI's Whisper model is fine-tuned for ASR, while neural MT and TTS architectures are optimised for translation and speech synthesis. Data preprocessing, feature extraction, and model evaluation are key steps in the methodology.

This study is grounded in theories of multilingual NLP, particularly transfer learning and fine-tuning (Houlsby et al., 2019). These theories suggest that pre-trained models on large datasets can be adapted to low-resource languages with task-specific data (Sanni, 2021). Advances in tonal recognition and speech synthesis provide the theoretical framework for addressing Hokkien's linguistic complexities.

The study is constrained by the availability and quality of datasets, which may not fully represent the linguistic diversity of Hokkien. Additionally, tonal accuracy in speech recognition remains a challenge, potentially impacting the system's effectiveness in preserving Hokkien's unique characteristics. As a result, the dataset used is focused on Taiwanese Hokkien, where the training dataset is the most abundant out of the other different variations of the Hokkien dialect.

To overcome the scarcity of Hokkien-to-English datasets, Chinese is chosen as an intermediate language for the transcribed Hokkien to be translated into, using machine translation, before translating into English. As there is an abundance of Chinese-to-English datasets, as well as a sufficient amount of Hokkien-to-Chinese datasets. Using a high-resource language as an intermediate language during machine translation can reduce the requirement for a training dataset as no dedicated Hokkien-to-English datasets is needed.

The system integrates three essential components: automatic speech recognition (ASR) to transcribe spoken Hokkien into text, machine translation (MT) to convert the Hokkien text into English using Chinese as an intermediate language, and text-to-speech (TTS) to synthesize natural-sounding speech in English, and vice versa. The ASR module is built on OpenAI's Whisper mode1, which serves as its foundation, while advanced neural architectures power the MT and TTS components. The system is fine-tuned on Hokkien datasets to ensure linguistic and contextual accuracy.

3. PRELIMINARY ANALYSIS

Preliminary evaluation suggest promising outcomes, as the performance of the Machine Translation model improves substantially in the first round of fine-tuning. However, further attempts to fine-tune the model using the same dataset yield minimal improvement, showing a need for a more diverse dataset. Beyond its technical contributions, this project is vital to preserving Hokkien's linguistic heritage and ensuring its accessibility for future generations (Kumar et al., n.d.).

By addressing the technological gap for Hokkien, this project aims to advance speech-to- speech translation for low-resource languages, contributing to both linguistic preservation and cross-lingual communication.

4. CONCLUSION

By exploring one possible method to develop a speech-to-speech translation model for low-resource languages, using Hokkien as our current focus, it is proven that by identifying intermediate, high-resource languages ensure that there are sufficient datasets to train the model to achieve high accuracy. This greatly reduces the requirements to achieve speech-to-speech translation of low-resourced languages in terms of dataset size, dataset quality, as well as available models. This shows a possibility for such method to be applied to other low-resourced languages to develop similar speech-to-speech models in pursuit of preserving their linguistic culture.

REFERENCES

Bahdanau, D., Cho, K., & Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. *arXiv (Cornell University)*.

Houlsby, N., Giurgiu, A., Jastrzebski, S., Morrone, B., De Laroussilhe, Q., Gesmundo, A., ... & Gelly, S. (2019, May). Parameter-efficient transfer learning for NLP. In *International conference on machine learning* (pp. 2790-2799). PMLR..

Klatt, D. H. (1987). Review of text-to-speech conversion for English. *The Journal of the Acoustical Society of America*.

Kumar, L. A., Renuka, D. K., Chakravarthi, B. R., & Mandl, T. (n.d.). *Automatic Speech Recognition and Translation for Low Resource Languages*. John Wiley & Sons.

Sanni, B. (2021). Transfer Learning in NLP: Designing Scalable Solutions to Address Low-Resource Language Challenges in Real-World Applications.

Sheu, J. S., Ahmad, A., & Xu, S. Y. (2024, July). Taiwanese Hokkien in AI: Challenges, Approaches, and Language Modeling. In *2024 International Conference on Consumer Electronics-Taiwan (ICCE-Taiwan)*.

Yu, D., & Deng, L. (n.d.). *Automatic Speech Recognition: A Deep Learning Approach*. Springer.

N. Law, G. Wong, E. Woo, & G. Jiang (Eds.) (2025). *Conference Proceedings of CITERS 2025.*
Hong Kong: The University of Hong Kong.

Enhancing Project Writing with GenAI in Higher Education: A Case Study of Two University Science Courses

Man Kin Leung[1], Marcus Ho-Yin Wong[1]
[12] The Hong Kong Polytechnic University, Hong Kong, China
[*] adam.leung@polyu.edu.hk

Abstract: This paper discusses the implementation of generative artificial intelligence (GenAI) in project assignments of two university courses. By analyzing students' projects written with the assistance of GenAI, the competencies required for effectively applying these technologies for project writing are evaluated. The characteristics and quality of the outcomes are compared to the previous works without using GenAI, which informs the modification of pedagogy and assessment design in STEM courses incorporating project-based learning.
Keywords: generative artificial intelligence (GenAI), project-based learning, technology-enhanced learning, higher education, STEM education

1. BACKGROUND

The emergence of generative artificial intelligence (GenAI) has significantly changed tertiary education. GenAI technology can create content in different forms and contexts in response to the prompts input by the user. This feature enables GenAI to assist with academic research and writing. As a result, many undergraduate students are eager to use GenAI in their course projects. This trend has caused universities to adjust their assessment policies to accommodate students' use of GenAI in assignments.

However, the feasibility and reliability of GenAI-assisted project writing in STEM education are questionable. Scientific writing is based on scientific facts with accuracy and consistency. With the reported issues of hallucination and inconsistency, use of GenAI in science projects must be examined carefully (Williams, 2024). There are also concerns about students over-relying on the use of GenAI without contributing their own ideas, or committing plagiarism without the acknowledgement of GenAI usage, etc. (Kasneci et al, 2023). In light of these concerns, it is necessary for educators to understand the competencies required for students to create high-quality projects using GenAI, while maintaining academic integrity and achieving intended learning outcomes.

2. OBJECTIVES

This paper aims to explore the essential competencies students need to effectively utilize GenAI in tertiary-level scientific writing, based on the empirical data collected from two project-based university courses. The strength and weakness of projects written with the use of GenAI will be evaluated. Additionally, the paper seeks to inform pedagogy and assessment design to foster students' capacity to use GenAI for project-based learning.

3. CONCEPTUAL FRAMEWORK

This study can be linked to the conceptual frameworks of project-based learning (PBL) and technology-enhanced learning (TEL). PBL emphasizes using project as both assessment and learning activities to enhance students' creativity, motivating them to apply their knowledge in real-world situations. PBL is particularly effective in higher education to foster students' learning and thinking skills (Balleisen et al, 2023). Students' usage of GenAI in writing can be evaluated in a continuum model of human-AI collaboration (Rowland, 2023).

TEL focuses on how technology can be integrated into teaching and learning activities to enhance students' learning (Kirkwood & Price, 2013). The implementation of GenAI and the development of new pedagogies can be evaluated through the lens of TEL. To achieve effective learning with new technologies, digital literacy of students is essential. This refers not only to the functional skills of using GenAI tools but also the ability to critically access the obtained information and convert it into own knowledge (Buckingham, 2013).

4. METHOD

The feasibility and effectiveness of implementing GenAI vary greatly among different disciplines and subjects. This involves pedagogy design, course context, teachers' perceptions of GenAI and students' GenAI literacy. Using case study approach, a course with such complexity can be regarded as a case, allowing the collection of rich, contextualized data for comprehensive analysis (Yazan, 2015).

In this study, two STEM-themed courses conducted by the authors at the Hong Kong Polytechnic University are selected as cases. "FSN4424 Food Product Development" is an elective subject for food science undergraduates, serving as a preparatory platform for their final-year projects. "AMA1D04 Understanding Social Conflicts by Game Theory" is a general education subject available for undergraduates from any disciplines. The enrollment numbers for Semester 1 of 2024/25 are 19 and 75 respectively. In both courses, students are required to submit a project assignment, with the use of GenAI allowed. They have received training on the techniques and integrities of using GenAI in project research and writing. Data collection and analysis are based on student surveys and project reports. Projects from previous years before the implementation of GenAI are used for comparison.

5. RESULTS AND SIGNIFICANCE

With the use of GenAI, students showed increased motivation and confidence to explore innovative ideas and tackle more complex projects. GenAI facilitated more comprehensive academic and marketing research, enabling students to gather extensive data that supports their projects. The quality of written English in reports improved substantially. Students could also create visual aids or even programs to illustrate their idea.

However, some drawbacks were identified in this study. A few projects relied heavily on AI-generated text, which was not adequately modified or tailored to the project's specific context. Some projects failed to review and validate references, thereby undermining the credibility of their findings. These observations highlight the need for a balanced approach to using GenAI, ensuring that students critically engage with GenAI outputs and maintain academic integrity.

Based on some high-quality projects and their reports on GenAI usage, students with high GenAI literacy are able to integrate their original idea and suggestions generated by GenAI to decide a coherent, authentic project topic. Their research and writing process leverage the strengths of both human and GenAI works. Such collaboration requires skills in critical thinking and decision-making, grounded in their knowledge in the course context.

ACKNOWLEDGEMENT

This study is a result of the project titled "Enhancing Teaching and Learning Experience with AI Technology", sponsored by the University Grants Committee's Fund for Innovative- Technology-in-Education (FITE).

REFERENCES

Balleisen, E. J., Howes, L., & Wibbels, E. (2023). The impact of applied project-based learning on undergraduate student development. *Higher Education*, *87*(4), 1141-1156. https://doi.org/10.1007/s10734-023-01057-1

Buckingham, D. (2015). Defining digital literacy-What do young people need to know about digital media? *Nordic Journal of Digital Literacy, 10*(Jubileumsnummer), 21-35.

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*. https://doi.org/10.1016/j.lindif.2023.102274

Kirkwood, A., & Price, L. (2013). Technology-enhanced learning and teaching in higher education: what is 'enhanced' and how do we know? A critical literature review. *Learning, Media and Technology*, *39*(1), 6-36. https://doi.org/10.1080/17439884.2013.770404

Rowland, D. R. (2023). Two frameworks to guide discussions around levels of acceptable use of generative AI in student academic research and writing. *Journal of Academic Language and Learning, 17*(1), T31-T69. https://journal.aall.org.au/index.php/jall/article/view/915

Williams, A. (2024). Comparison of generative AI performance on undergraduate and postgraduate written assessments in the biomedical sciences. *International Journal of Educational Technology in Higher Education*, *21*(1). https://doi.org/10.1186/s41239-024-00485-y

Yazan, B. (2015). Three approaches to case study methods in education: Yin, Merriam, and Stake. *The qualitative report, 20*(2), 134-152.

Collaborative Language Learning Through Technology: A Systematic Review

Zehao Li[1*], Zeekin Ziqian Zhou[2]
[1]Teacher College, Columbia University, U.S.A.
[2]School of English, The University of Hong Kong, Hong Kong, China
[*] zl3323@tc.columbia.edu

Abstract: This paper reviews instructors' technology adoption in foreign or second language classes, focusing on how technological tools have enhanced collaborative learning in the decade of 2014-2024. Based on an in-depth review of 52 empirical studies based on the PRISMA guidelines, this paper presents the the most commonly used technological tools that are either synchronous (i.e., video meeting, social media, and gamified platforms) or asynchronous (e.g., cloud literacy platforms and learning-management systems. Selected articles indicated those tools generally have a positive impact on students' learning.
Keywords: technology-enhanced learning; computer-assisted language; learning; emerging technology in education

1. INTRODUCTION

Information technology has been introduced to teaching and students' learning, communication, and collaboration since the 1970s (Harasim, 2000; Wen et al., 2012). Numerous terms related to the use of technology in collaborative learning have emerged, including "telecollaborative" (O'Dowd, 2010), "computer-supported collaborative learning" (Ernest et al., 2012), and "technology-enhanced collaborative language learning" (TECLL) (Su & Zou, 2020). TECLL is an cross-disciplinary field in which language educators draw theoretical and practical inspirations from computer science, education, and linguistics into classroom teaching and learning. The TECLL pedagogy requires teachers and students to interact and collaborate using technologies either for the digitalisation of communication, co-editing, review and feedback-giving, or co-reading and discussion.

However, most previous studies were conducted in the broader context of computer-assisted language learning or collaborative learning, predominantly proposing theoretical approaches or evaluating empirical implementations, lacking a systematic review focusing on the collaborative elements. Exceptionally, Su and Zou (2020) reviewed 40 articles on TECLL, whereas they emphasis mostly on teaching and learning theories, leaving the practical use and means of implementing technology in language pedagogy to be further reviewed.

To address this gap, the present study aims to systematically review articles on how they pedagogically adopt technology for TECLL, focusing on examining the features of the tech tools and their corresponding implications. The following questions guide the research:

1. What technological collaborative tools have been used for TECLL?

2. How are these tools implemented within and/or beyond classroom learning?

3. How are the tools effective in students' learning in terms of collaboration and language skills?

2. METHODOLOGY

This study draws upon the generic steps proposed by Templier and Paré (2015) and the flow diagram in the PRISMA statement (Page et al., 2021) to form the method flow demonstrated below in Fig. 1.

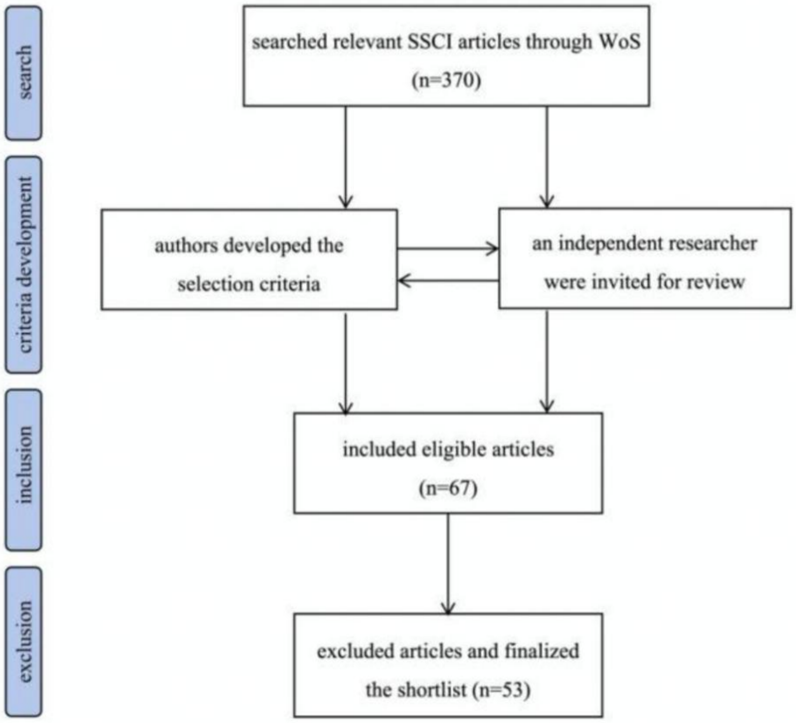


Fig. 1. Search and selection procedure of the articles.

2.1 Selection Criteria

The articles were collected from the Web of Science (WoS) Core Collection, which is a recognized academic pool of high-quality papers. The present study included articles if they were aligned with all of the following inclusion requirements:

- The study must be empirical.
- The study must use technology to enhance collaborative learning.
- The study must aim to improve language learning through technology.
- The study must provide clear information regarding the learning design.
- The study must provide a clear explanation of the technology used or developed.

To further refine the scope and ensure the quality of selected studies, articles were excluded if they exhibited any of the following characteristics, even if they seemingly met the basic inclusion criteria:

- The designed collaborative activity could occur independently without technology.
- The focus was on the development of the technology rather than its application to

enhance learning.
● The study mainly focuses on other perspectives rather than TECLL, such as learner identity or professional teacher development.

We conducted a validation check of the screened articles by searching a sample of 20 of them in ERIC and Scopus, which helps verify that the paper collection from the WoS does not involve sources that are excluded by high-impact journal index.

3. TECHNOLOGY APPLIED IN COLLABORATIVE LANGUAGE LEARNING

In this study, 52 articles were selected and thoroughly reviewed based on the above-mentioned criteria. The selected papers were categorized according to the features of the technology, which were finally divided into seven categories (see Table 1).

Category	Subcategory		Technology	Articles
Synchronous collaborative tools	Social media		Line	Chang and Lu (2018), Wu et al. (2017)
			CoveritLive	Zheng and Warschauer (2019)
			aLF e-learning platform	Talaván et al. (2016)
			Self-developed	Barrett et al. (2022), Huang (2015), Shadiev et al. (2018)
			Qzone	Xu and Yu (2018)
			WhatsApp	Andujar (2016), Avci and Adiguzel (2017)
			WeChat	Chen and Du (2022), Dai and Wu (2022), Jiang et al. (2021), Wang and Jiang (2024)
			Tencent QQ	Jiang and Eslami (2022)
	Video-meeting tools		Skype	Dooly and Sadler (2016), Kato et al. (2023)
			Zoom	Aubrey (2022)
	Gamified learning tools	Digital gameplay	MMORPG	Ng et al. (2022)
			Ragnarok Online	Reinders and Wattana (2014)
			ChronoOps	Sydorenko et al. (2019)
			Microsoft Kinects	Wang et al. (2019)
			Meet-Me	Yamazaki (2018)
			Self-developed	Chu et al. (2019), Liu (2022)
		Single-display groupware	Digital Mysteries	Lin et al. (2016)
			Self-developed	Calderón et al. (2016)

Asynchronous collaborative tools	Cloud literacy platforms	Editing tools	Google Docs	González-Cruz et al. (2022), Hoang and Hoang (2022), Hsu (2022), Liu et al. (2014), Yeh and Chen (2019), Zou and Xie (2019)
			Google Scribbles	Wen et al. (2015)
			Wiki	Chew and Ding (2014), Hsieh (2017), Hsu and Lo (2018), Li and Chu (2018), Such (2021), Wang (2014, 2015)
			Ebeam	Teng (2021)
			NetSupport	Gleason (2014)
			Poetry Zone	Lan et al. (2015)
		Annotatio n tools	Perusall	Kohnke and Har (2022), Zhang and Li (2023)
			Self-developed	Manabe et al. (2021), Zhang et al. (2019)
	Learning-management systems		Edmodo	Adhami and Taghizadeh (2022)
			Blackboard	Angelova and Zhao (2016)
			WikiTalki	Ko and Lim (2022)
			Self-developed	Yang (2016)

The synchronous communication tools include those for video meeting, social media communication, and learning gamification. Video meeting tools apply multimodal sources to allow students to conduct online discussions without the restriction of space and distance. Social media, similarly, facilitates online communication, whereas they support mainly message exchange instead of real-time talk. Although some social media platforms involve the video-call function, students using these platforms adopt multiple functions (e.g., text and audio exchange, blog sharing, use of memes, etc.) rather than keep having online meetings (Chen & Du, 2022). The technologies for learning gamification involve teachers either conducting digital virtual games or letting students collaborate on a single-display groupware to complete game tasks. Gamification has been used in education to a larger extent than its application in most other industries (Hamari et al., 2014).

4. LEARNING DESIGNS TO ADOPT THE TECHNOLOGIES

While the synchronous tools are mainly adopted for students' communication platform constructions, the asynchronous tools are incorporated mostly for literacy skills. As synchronous tools, social media are primarily used for after-class communication among students while they are collaborating. In contrast, video-meeting tools are mostly adopted in class for students with special a gap to talk and collaborate on oral tasks. Gamified learning design is usually task-based as well, which either aims at enhancing students' communication skills or to practice their knowledge items by co-finishing (e.g.) grammar or vocabulary ventures (Ng et al., 2022). Unlike the synchronous tools that facilitate instant communication, the asynchronous tools usually aim at more sustainable, long-lasting collaborations. Teachers with asynchronous tools usually implement the adoption of these tools throughout a whole course that lasts at least one semester or with a number of class series, in contrast to many synchronous tools that last only for one class.

5. LEARNING EFFECT OF TECLL

As aligned with the pedagogical design in the last section, synchronous tools benefit mostly students' communicative skills, and asynchronous tools extend primarily students' knowledge of the target language, such as grammar and vocabulary, adding writing skills as developed by the use of cloud literacy platforms. It is interestingly found that while video-meeting tools improve students' oral skills, social media benefits their use of multiple linguistic sources, or language repertoire, while communicating. It makes sense since video meeting requires students to talk orally, and social media allows students to draw on multiple language resources, not only text, audio, but also cross-cultural memes, pictures, etc., which thus supports students' practice of translanguaging (Li, 2022), using the target language and its related elements not only as a code but a flexible way of meaning-making. In terms of gamification, it is found that the enhancement of learning engagement is larger than that of learning effect, which means gamified pedagogy is mainly to engage students rather than improve teaching efficiency. Unlike synchronous TECLL studies that usually compare technology incorporation with traditional pedagogy, studies of asynchronous tools usually do not involve parallel comparisons since they do not aim at enhancing efficiency but extend the class to remote contexts. Thus, the summary of the effect of these tools is without comparison, but overall effective; they help realise new teaching and learning forms.

6. DISCUSSION

The systematic review reveals a transformation in the field of collaborative language learning technology over the past decade. The review indicates that the success of TECLL fundamentally depends on how well technological tools are integrated with pedagogical objectives and instructional design, rather than merely the presence of technology. A comparison of these technology categories found differences in usage frequency and learning modes and outcomes influenced by their design characteristics, which will be further explained in the presentation and the full paper. The significance of this study lies in providing a comprehensive overview to practitioners who are thinking about integrating technologies to enhance collaborative language learning.

REFERENCES (THE 53 REVIEWED ARTICLES ARE NOT LISTED DUE TO THE PAGE LIMIT)

Chen, C., & Du, X. (2022). Teaching and learning Chinese as a foreign language through intercultural online collaborative projects. *The Asia-Pacific Education Researcher*, 31, 123–135. https://doi.org/10.1007/s40299-020-00543-9

Ernest, P., Guitert Catasús, M., Hampel, R., Heiser, S., Hopkins, J., Murphy, L., & Stickler, U. (2012). Online teacher development: collaborating in a virtual learning environment. *Computer Assisted Language Learning*, *26*(4), 311–333. https://doi.org/10.1080/09588221.2012.667814

Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does Gamification Work? — A literature review of empirical studies on gamification. In the Proceedings of the *47th Hawaii 5 International Conference on System Science*. 3025–34. https://doi.org/10.1109/HICSS.2014.377

Harasim, L. (2000). Shift happens: online education as a new paradigm in learning. *The Internet and Higher Education, 3*(1–2), 41–61. https://doi.org/10.1016/S1096-7516(00)00032-4

Li, W. (2022). Translanguaging as method. *Research Methods in Applied Linguistics*, *1*(3), 100026. https://doi.org/10.1016/j.rmal.2022.100026

Ng, L. L., Azizie, R. S., & Chew, S. Y. (2022). Factors influencing ESL players' use of vocabulary learning strategies in massively multiplayer online role-playing games

(MMORPG). *The Asia-Pacific Education Researcher, 31*(4), 369–381. https://doi.org/10.1007/s40299-021-00578-6

O'Dowd, R. (2010). Intercultural communicative competence through telecollaboration. In J. Jackson (Eds.), *The Routledge handbook of language and intercultural communication* (pp. 340–356). Routledge. https://doi.org/10.4324/9780203805640.ch21

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*, 89. https://doi.org/10.1186/s13643-021-01626-4

Su, F., & Zou, D. (2020). Technology-enhanced collaborative language learning: theoretical foundations, technologies, and implications. *Computer Assisted Language Learning, 35*(8), 1754–1788. https://doi.org/10.1080/09588221.2020.1831545

Templier, M., & Paré, G. (2015). A framework for guiding and evaluating literature reviews. *Communications of the Association for Information Systems, 37*, 112–137. https://doi.org/10.17705/1CAIS.03706

Wen, Y., Looi, C.-K., & Chen, W. (2012). Supporting teachers in designing CSCL activities: A case study of principle-based pedagogical patterns in networked second language classrooms. *Educational Technology & Society, 15*(2), 138–153.

N. Law, G. Wong, E. Woo, & G. Jiang (Eds.) (2025). *Conference Proceedings of CITERS 2025.*
Hong Kong: The University of Hong Kong.

Analysis of Classroom Dialogue Sequence Characteristics in Primary School Science Lessons Under the Framework of Scientific Core Literacy

Xueni Liu[1], Junwei Chen[2*], Minjie Zhu[1], Xilin Wu[1] & Sijing Yu[1]
[1]School of Education Information Technology, South China Normal University, China
[2]School of Psychology, South China Normal University, China
[*] 1029903092@qq.com

Abstract: Classroom dialogue is vital for students' knowledge construction, yet its effectiveness in primary science classrooms remains limited. This study employs learning analytics to analyze 758 minutes of exemplary lessons, examining dialogue sequence patterns and identifying high-quality interactions. The findings reveal the deep-seated principles of effective classroom dialogue in primary science education and indicate that teachers should adopt a student-centered approach, enhance teacher-student interaction, facilitate active classroom dialogue, and optimize learning feedback. This study provide valuable insights for improving primary science teaching and fostering students' knowledge construction.
Keywords: Science education; Classroom dialogue; Knowledge construction; Sequence mining

1. INTRODUCTION

Amidst the new wave of technological revolutions and artificial intelligence innovations, science education profoundly influences national development. The Compulsory Education Science Curriculum Standards (2022 Edition) of China (hereinafter referred to as the "New Curriculum Standards") emphasize that science education aims to cultivate students' core literacy. Classroom dialogue in science education carries rich information and serves as an effective medium for analyzing teacher-student interactions. High-quality science classroom dialogue fosters students' knowledge construction and enhances scientific literacy (Liu & Wen, 2015). As research on classroom dialogue advances, effective dialogue has been identified as a crucial element in avoiding mere knowledge transmission. However, science teachers tend to adopt authoritative dialogue approaches, neglecting students' agency in the classroom (Zhou et al., 2018). Existing research on classroom dialogue primarily focuses on discourse classification or coding frameworks, often employing descriptive statistics and qualitative analysis, with limited integration of these methods (Howe & Abedin, 2013). Moreover, previous studies have typically analyzed only a few lesson examples using manual coding, making it challenging to capture patterns in classroom interaction. Therefore, enhancing the specificity and professionalism of science teachers' instructional dialogue to improve teacher-student interaction effectiveness warrants further exploration.

Learning analytics technology, which integrates educational data collection, processing, information extraction, and teacher decision support, can assist educators in improving instructional discourse precision (Wei, 2013). Hence, this study employs learning analytics technology to analyze teacher-student dialogue in science classrooms, extracting dialogue characteristics and structural patterns to summarize effective teacher-student dialogue pathways and propose strategies for conducting effective science classroom dialogue.

2. LITERATURE REVIEW

2.1 Characteristics of Primary Science Classrooms Under the Framework of Scientific Core Literacy

Primary science classrooms emphasize student-centered learning, engaging students in scientific inquiry through dialogue-based discussions. The New Curriculum Standards advocate for increased inquiry-based activities and the integrated use of heuristic, interactive, and exploratory teaching approaches to enhance students' scientific core literacy (Shi, 2022). In this context, the teacher's role shifts from a knowledge transmitter to a facilitator of students' knowledge construction, transforming science instruction from knowledge control to inquiry support. This approach underscores the importance of teacher-student interactions and students' knowledge construction (Hu et al., 2021), fully leveraging student agency.

2.2 Research on Science Classroom Dialogue Analysis

Teacher-student dialogue, as a core classroom element, has attracted significant scholarly attention, with research focusing on dialogue coding systems, interaction patterns, and dialogue sequences. Various dialogue coding frameworks have been applied to science classrooms, including Flanders' Interaction Analysis Categories, S-T Analysis Framework, and customized coding systems tailored for science education, such as the Educational Dialogue Analysis Teacher Scheme and Knowledge Construction Coding Framework. Common dialogue structures and patterns include "Initiation-Response-Feedback" (IRF), teacher monologue, peer discussion, and authoritative discourse (Viiri & Saari, 2006). Studies analyzing dialogue sequences often identify significant behavior patterns through frequency statistics, while others use combined behavior sequences to reveal teacher-student interaction effects (Jiang et al., 2019). Existing research frequently integrates classroom dialogue models and sequence analysis to clarify specific dialogue characteristics and patterns, designing dialogue coding schemes based on discourse structure, discourse types, and interaction models.

2.3 Application of Learning Analytics Technology in Science Classrooms

Learning analytics technology facilitates the collection, analysis, and reporting of learner-environment interaction data. Researchers have begun applying this technology to science education, such as using classroom dialogue analysis to guide teaching behavior (Calcagni et al., 2023), understanding student learning behaviors and cognitive processes, employing interaction analysis to enhance group discussions, and leveraging convolutional neural networks with long short-term memory structures to extract online discussion data for predicting students' critical thinking levels (Calcagni et al., 2023).

2.4 Research Questions

Primary science classrooms exhibit strong interactivity and emphasize student knowledge construction. Despite extensive research on science classroom dialogue, existing coding frameworks rarely address the impact of teacher-student dialogue on students' knowledge construction at the primary level. Additionally, the extent to which teachers connect prior knowledge to new scientific concepts remains unclear. Furthermore, the use of learning analytics technology in primary science classroom dialogue research is relatively underexplored. Therefore, this study aims to design a knowledge construction coding framework for primary science classrooms, incorporating learning analytics technology to extract teacher-student dialogue features and structural patterns, ultimately summarizing effective teacher-student dialogue pathways.

The study seeks to address the following questions:

1. What are the characteristics of teacher-student dialogue in exemplary primary science lessons?
2. What significant dialogue sequences exist in exemplary primary science lessons?
3. How does teacher-student dialogue impact students' knowledge construction in exemplary primary science lessons?

3. METHOD

3.1 Context and Participants

The study uses video recordings from the "One Teacher, One Excellent Lesson, One Famous Teacher" program of the Ministry of Education as data sources. Lesson selection followed these criteria: (1) High-rated, widely viewed, and recently published national-level exemplary science lessons, as they are widely recognized by teachers and experts; (2) Lessons with clear teacher-student dialogue and smooth instructional delivery, ensuring a comprehensive representation of real science classrooms. An initial selection yielded 34 lessons spanning primary and secondary levels. Based on dialogue sequence analysis dimensions related to teacher-student dialogue and knowledge construction, 18 primary science lessons meeting the criteria were selected, yielding 758 minutes of video data. After excluding irrelevant segments, the final dataset comprised classroom dialogues averaging 43 minutes per lesson.

3.2 Measures

The study employs an IRF discourse structure (Lyle, 2008) to classify teacher- student dialogues into Initiation-Response-Feedback triadic interactions. First, the study adapts a teacher discourse coding framework by integrating the T-SEDA (The Teacher Scheme for Educational Dialogue Analysis) framework (Hennessy et al., 2021) and the knowledge construction dimension dialogue coding framework proposed by Li et al. (2022). Second, a student discourse coding framework is adapted based on Zhang et al. (2020), who utilized ETIAS coding in smart classroom teaching behavior research, and the knowledge construction dialogue text analysis framework of high-quality courses (Li, 2014). Finally, to comprehensively analyze teacher- student dialogue in science classrooms, the study integrates both teacher and student discourse coding frameworks through pre-coding, refinement, and consolidation, ultimately forming the science classroom knowledge construction coding framework (Table 1).

Table 1: Knowledge Construction Coding Framework for Science Classrooms

Dialogue Structure Type	Revised Category	Description
Inquiry	Initiating Inquiry (II)	The teacher prompts students to express their thoughts, opinions, emotions, and insights about scientific phenomena and ideas, helping them better understand and master scientific knowledge.
	Guiding Analysis (GA)	The teacher guides students in analyzing scientific problems in

		depth, helping them grasp the essence of scientific concepts and reasoning, thereby improving scientific literacy and problem-solving skills.
	Transfer and Inference (TI)	The teacher supports students in developing innovative thinking and applying knowledge transfer, enhancing their ability to solve scientific problems creatively. This approach helps students utilize and expand their scientific thinking and innovation abilities based on fundamental scientific knowledge.
	Student-Initiated Questioning (SQ)	Students actively pose questions to the teacher, such as introducing a new inquiry or responding to content. This allows students to engage with the lesson by expressing their thoughts and curiosity.
Response	Mechanical Response (MR)	Students provide passive, simple, and mechanical answers to teachers' questions or dialogues. In such cases, teachers can guide students to think more deeply to enhance their understanding of scientific knowledge and problems.
	Cognitive Response (CR)	Students actively answer teachers' questions based on their existing scientific knowledge and demonstrate logical reasoning and explanation skills. This response type allows students to use their scientific knowledge and critical thinking abilities in responding.
	Inferential Response (IR)	Students think and analyze before answering teachers' questions. Their responses are based on prior scientific knowledge, logical reasoning, judgment, and interpretation, allowing them to explain their ideas through logical inference.
	Innovative Response (IR)	Students reflect on the teachers' questions and respond based on their own innovative thoughts and understanding. This response type

		enables students to leverage their creative thinking and cognitive abilities.
Feedback	Knowledge Explanation (KE)	The teacher conveys fundamental scientific knowledge and concepts, helping students learn basic scientific methods and principles to grasp essential elements of scientific knowledge.
	Connection and Synthesis (CS)	The teacher synthesizes and summarizes scientific knowledge by drawing connections and comparisons, enabling students to understand the fundamental principles behind scientific phenomena.
	Responsive Expansion (RE)	When a student's response is acknowledged by the teacher or peers, the teacher encourages further discussion, prompting students to engage in more interactive and expansive dialogues.
	Affirmation and Encouragement (AE)	The teacher provides targeted guidance and support during the learning process, encouraging students based on the lesson's progress and their learning stages.
	Questioning and Reflection (QR)	The teacher critically evaluates students' responses by analyzing and providing feedback. Additionally, when students challenge or question responses, the teacher fosters active participation and reflective thinking.
	Guidance and Support (GS)	The teacher provides structured guidance based on the lesson plan and students' learning progress, offering instructional support and motivation.

3.3 Data Processing

Finally, to comprehensively analyze teacher- student dialogue in science classrooms, the study integrates both teacher and student discourse coding frameworks through pre-coding, refinement, and consolidation, ultimately forming the science classroom knowledge construction coding framework (Table 1). This study adopts a data analysis approach consisting of pattern analysis, pre-coding, consistency testing, formal coding, and sequence analysis. Based on the dialogue structure proposed by Lemke (1983), teacher-student dialogues in science classrooms are classified into four categories: segments, sequences, transitions, and speech moves. Segments refer to all dialogue content produced by different dialogue participants within a specific activity; sequences describe the transformation of one piece of information into a related one; transitions occur between two speakers and consist of two speech moves; and speech moves represent the "smallest unit" of dialogue.

The data processing procedure is as follows: First, the Observed Known on Classroom software was used in conjunction with the teacher-student dialogue coding framework to encode all dialogue texts from each science lesson line by line, successfully obtaining a total of 5,447 speech move-level dialogue data points. Next, these data were sequentially encoded to extract transition-level data in science classrooms. Finally, the General Sequential Querier (GSEQ) software was employed to analyze the transition data, determining whether classroom dialogue sequences exhibited statistical significance and ultimately producing a frequency analysis of dialogue interaction pattern transitions relevant to this study.

4. RESULTS AND DISCUSSION

4.1 Statistical Analysis of Dialogue Patterns

This study analyzes primary school science classroom dialogue patterns from three perspectives: the number and roles of dialogue participants, the structural types of dialogue based on classroom dialogue models, and the distribution of dialogue dominance.

4.2 Consistency Testing of Coded Data

The coding process was independently conducted by two researchers, one serving as the primary coder and the other as the secondary coder. First, both researchers familiarized themselves with the coding framework for knowledge construction dialogue in science classrooms. Second, they coded two selected classroom videos. Third, the SPSS software was used to perform a Kappa consistency test on the 304 coded data entries obtained by both coders. The final Kappa value was 0.857, indicating a high level of data consistency.

4.3 Processing and Analysis of Coded Data

The study employed learning analytics software GSEQ 5.1 to analyze dialogue sequence data. First, the sequence data were imported into the software to generate a behavioral transition frequency table and an adjusted residual table for dialogue sequences to determine the statistical significance of each behavioral sequence. Finally, statistically significant sequences were classified into three categories (Table 2).

Table 2: Significant Sequences of Different Interaction Patterns in Exemplary Science Lessons

Type	Behavior Sequence (Frequency/Z-Value)
Teacher or Student Initiation	II→SQ (14/5.82), II→MR (239/27.00), II→CR (333/31.15), II→AE (77/10.94), II→IR (32/4.96), GA→GA (66/2.88), GA→AE (119/6.49), GA→CS (60/5.92), MR→RE (30/10.29), MR→IR (10/5.06), SQ→KE (14/11.47)
Teacher or Student Response	MR→II (79/2.96), MR→AE (72/6.05), CR→KE (42/2.50), CR→AE (168/16.90), CR→IR (12/3.15), IR→AE (41/6.15), IR→QR (16/9.19), IR→IR (14/6.65), IR→CS (13/4.55), IR→RE (11/6.02), IR→AE (31/7.42)
Teacher Feedback	KE→KE (66/7.27), KE→GA (56/3.27), KE→CS (37/2.97),

	KE→MR (58/12.66), CS→KE (52/6.30), CS→II (97/3.96), CS→RE (24/4.55), CS→IR (15/2.91), RE→KE (37/7.24), RE→CS (20/3.26), RE→RE (14/3.77), AE→II (234/6.87), AE→CS (71/4.46), AE→RE (34/2.25), AE→GS (154/2.46), QR→GA (22/4.37), QR→QR (7/3.88), GS→II (279/6.69), GS→GS (269/10.09)

5. DISCUSSION AND CONCLUSIONS

5.1 Dialogue Patterns in Exemplary Science Lessons Tend Toward Initiation Response-Feedback Structure

The results indicate that dialogue patterns in exemplary science lessons predominantly follow an Initiation-Response-Feedback (IRF) structure. Although this structure is common, in science teaching, it facilitates lesson progression through multiple rounds of teacher-student dialogue, fostering deeper engagement, breaking the closed nature of the classroom, and promoting students' knowledge construction (Wang & Yang, 2024). Therefore, teachers should adopt different dialogue approaches at various instructional stages—such as basic knowledge instruction and phenomenon inquiry—to guide students in developing a deeper understanding of scientific concepts.

5.2 Significant Dialogue Sequences in Science Classrooms Facilitate Knowledge Construction

As shown in Table 2, there are 41 statistically significant dialogue sequences with a residual value (Z) greater than 1.96, including 11 teacher- or student-initiated sequences, 11 response sequences, and 19 feedback sequences. Visualized through Gephi, these sequences provide insights into significant dialogue behaviors (Fig. 1-3).

Constructivist theory emphasizes the importance of social interaction, asserting that knowledge is constructed through engagement with the environment. Classroom dialogue fosters this process. Figure 1 illustrates that teacher-initiated questioning plays a dominant role, with significant sequences such as PE→SA, PE→MA, PE→CA, PE→RA, and PE→IA, demonstrating that prompting students to think and express their ideas supports knowledge construction. Additionally, sequences GA→GA, GA→R, GA→CI, MR→RE, and MR→IA indicate that guided and transfer-based instructional prompts effectively direct student responses. Notably, student-initiated questioning, though infrequent (SA→KT), leads to subsequent knowledge explanation, benefiting students' learning. Therefore, teachers should enhance student agency by providing more opportunities for questioning.

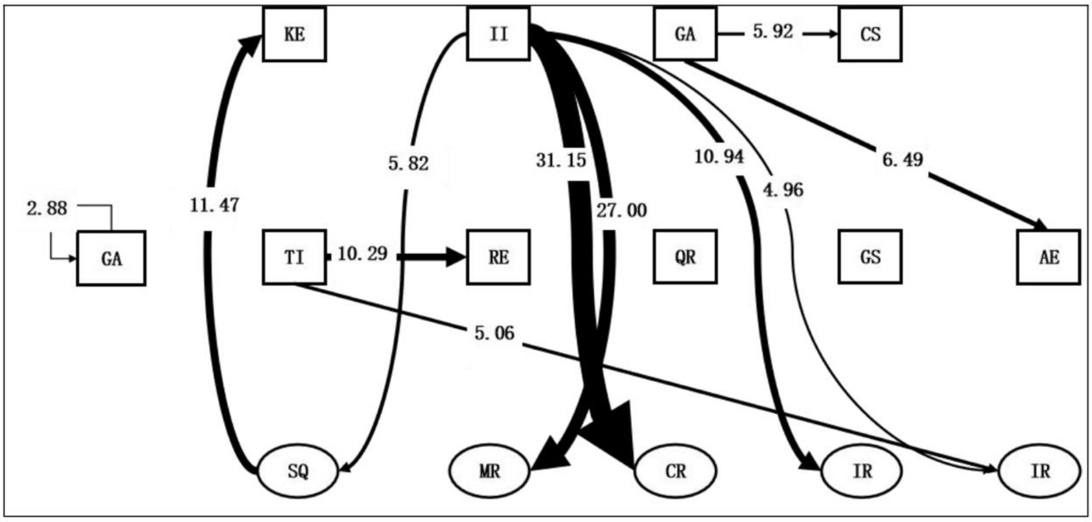


Fig. 1. Significant Teacher- or Student-Initiated Dialogue Sequences

Sequences MA→PE and MA→R suggest that when students provide mechanical responses, follow-up questioning or encouragement from teachers facilitates deeper learning (Fig. 2). Sequences CA→KT, CA→R, and CA→IA demonstrate that teachers encourage students to respond using their prior knowledge. Although inferential and innovative responses are less frequent (RA→R, RA→Q, RA→RA, IA→CI, IA→RE, IA→R), they prompt teachers to expand on students' ideas, summarize knowledge, and guide further inquiry.

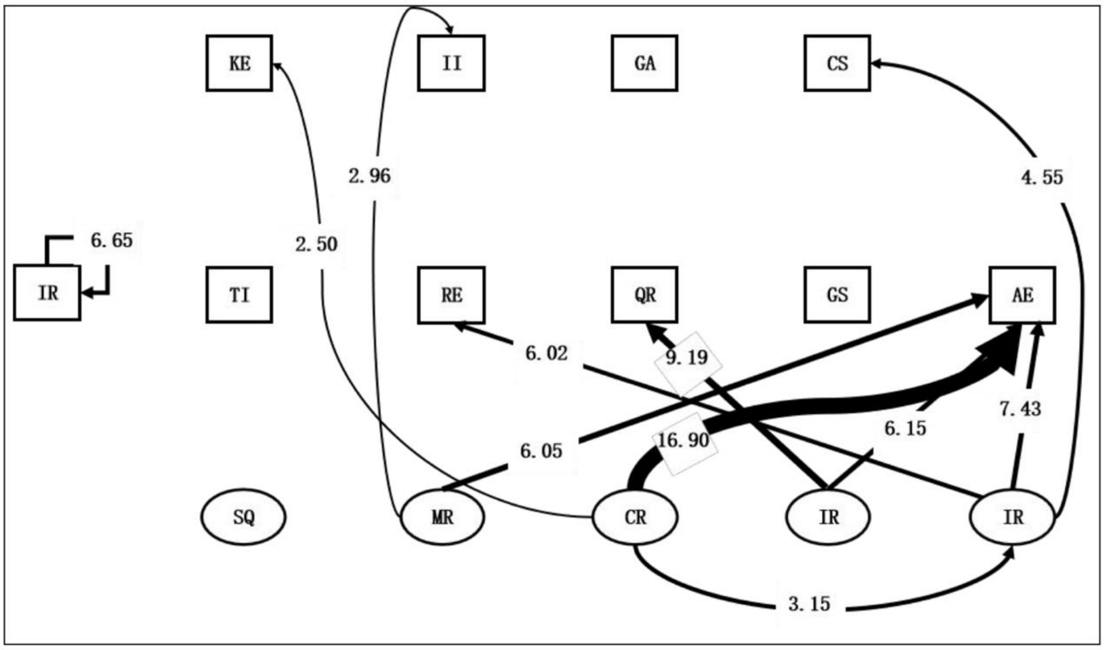


Fig. 2. Significant Response Sequences in Teacher-Student Dialogue

In teacher feedback sequences (Fig. 3), teachers employ diverse feedback strategies, emphasizing encouragement and knowledge elaboration. Significant sequences such as KT→KT, KT→GA, KT→CI, KT→MR show that teachers primarily provide instructional support and explanations to guide students in analysis and synthesis. CI→KT, CI→PE, CI→RE, CI→RA reveal that summarization, adopting student ideas, providing feedback, and reasoning support deep learning. Additionally, multiple feedback loops appear (RE→KT, RE→C, RE→RE, R→PE, R→CI, R→RE, R→G, Q→GA, Q→Q, G→PE, G→G), indicating that teachers use guidance, praise, and questioning strategies to foster reflection and discussion, creating a supportive classroom environment that aligns with the 2022 Compulsory Education Science Curriculum Standards.

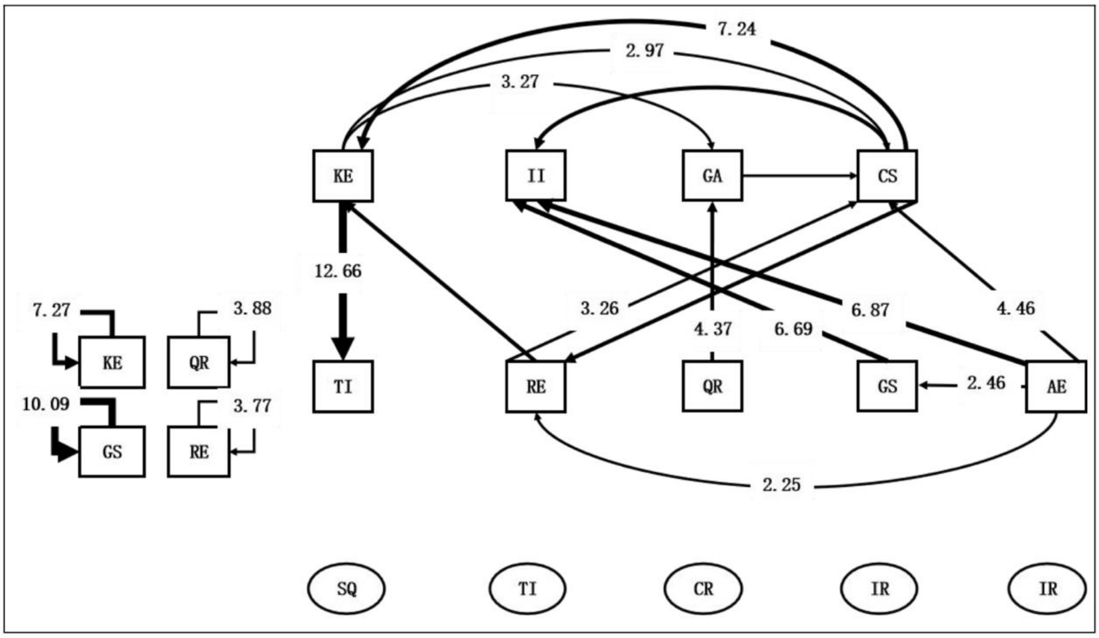


Fig. 3. Significant Teacher Feedback Sequences in Science Classrooms

5.3 Diverse Effective Sequences in Teacher-Initiated Dialogue

Teacher-initiated dialogue exhibits diverse and effective sequences, particularly in feedback adjustments based on student responses (Fig. 4). For example, cognitive, inferential, and innovative responses generate a broader range of feedback types, while affirmation and encouragement foster more extended dialogue exchanges. Knowledge construction is hierarchical and generative, and teacher discourse helps students form question chains, activate thinking processes, and stimulate knowledge production (Zhu et al., 2024). Therefore, in inquiry-based science teaching, teachers should not only affirm and encourage students' responses but also tailor their feedback based on the response type—especially promoting inferential reasoning and innovative thinking—to facilitate effective classroom dialogue and deeper knowledge construction.

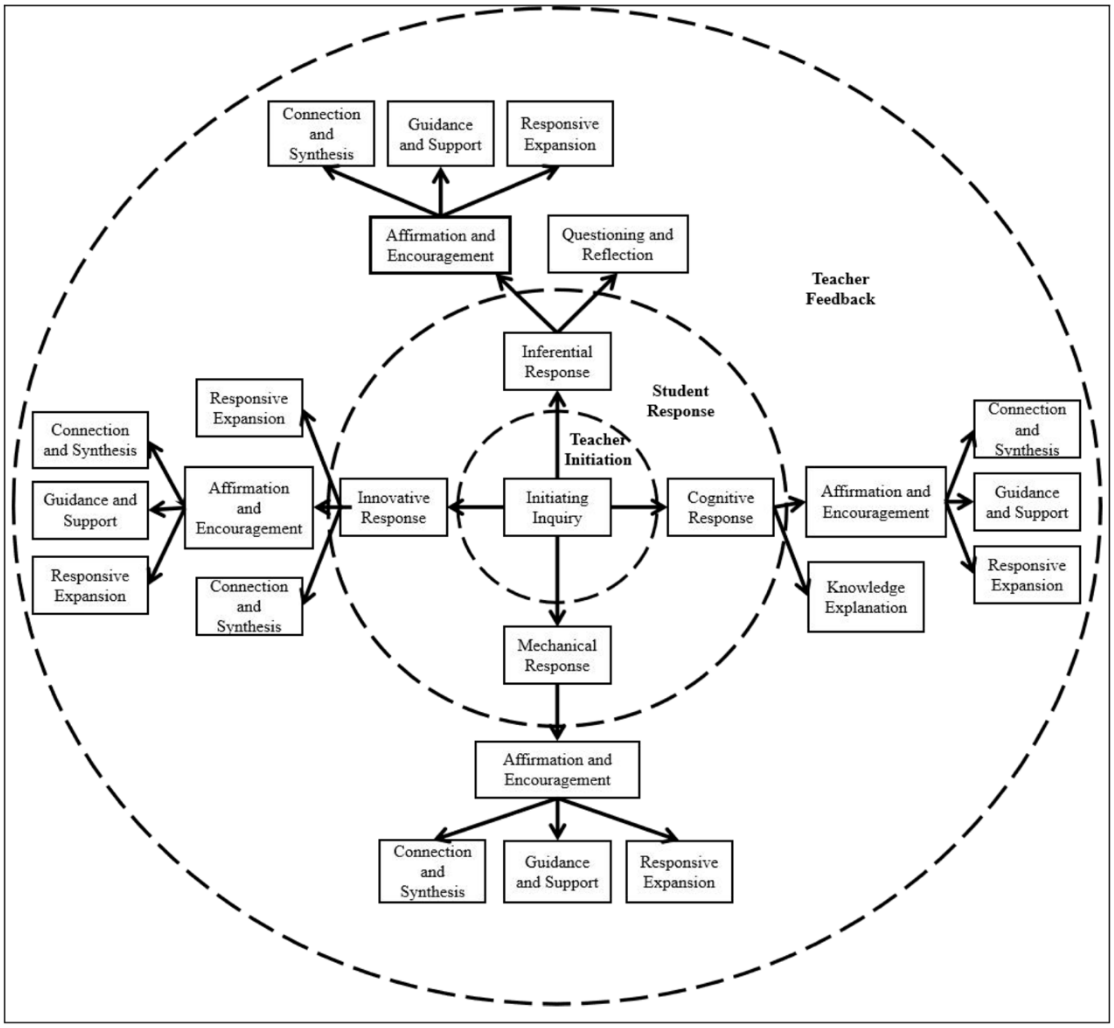


Fig. 4. Significant Teacher-Led Dialogue Pathways in Science Classrooms

5.4 Student-Initiated Dialogue Contributes to Deep Knowledge Construction

The analysis reveals that student-initiated dialogue is relatively rare, yet when students do initiate dialogue, the sequences are statistically significant and effective (Fig. 5). This process represents an adjustment of students' cognitive schemas, leading to higher-order knowledge construction (Zhong, 2006). During this process, teacher feedback plays a crucial role in shaping students' scientific literacy development. Therefore, when students raise questions, teachers should first analyze them in relation to classroom knowledge, followed by facilitating knowledge transfer, inferential reasoning, and synthesis to guide students toward new insights.

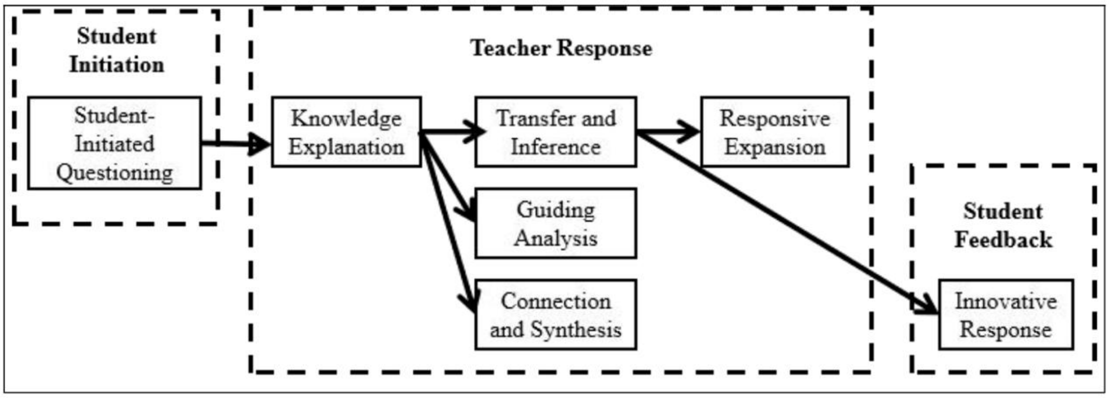


Fig. 5. Significant Student-Initiated Dialogue Pathways in Science Classrooms

6. IMPLICATIONS AND LIMITATIONS

In high-quality classrooms, science teachers' discourse strategies effectively facilitate students' knowledge construction. However, in typical primary school science classrooms, while teacher discourse demonstrates a high level of interactivity, it remains relatively insufficient in promoting students' conceptual development. Therefore, teachers should place greater emphasis on students' construction of scientific knowledge by designing structured and targeted open-ended question sequences and employing heuristic, inferential, and innovative questioning techniques. These approaches encourage students to engage in deep exploration and comprehension of scientific concepts. The focus should shift from merely pursuing "correct answers" and "correcting wrong responses" to "stimulating accurate understanding" and "correcting misconceptions". This transformation will help students progressively build a comprehensive conceptual framework of scientific knowledge, significantly enhancing their scientific core literacy.

REFERENCES

Calcagni, E., Ahmed, F., Trigo, A., & Kershner, R., & Hennessy, S. (2023). Developing dialogic classroom practices through supporting professional agency: Teachers' experiences of using the T-SEDA practitioner-led inquiry approach. *Teaching and Teacher Education*, *123*, 104067.

Calcagni, E., Ahmed, F., Trigo-Clapés, A. L., & et al. (2023). Developing dialogic classroom practices through supporting professional agency: Teachers' experiences of using the T-SEDA practitioner-led inquiry approach. *Teaching and Teacher Education*, *126*, 104067.

Hennessy, S., Kershner, R., Calcagni, E., & Ahmed, F. (2021). Supporting practitioner-led inquiry into classroom dialogue with a research-informed professional learning resource: A design-based approach. *Review of Education*, *9*(3), e3269.

Howe, C., & Abedin, M. (2013). Classroom dialogue: A systematic review across four decades of research. *Cambridge Journal of Education*, *43*(3), 325-356.

Hu, H., Li, M., & Yang, W. Y., et al. (2021). Progress and prospects of junior high school science classroom teaching reform—Based on video analysis of Shanghai junior high school science classrooms in 2006 and 2016. *Basic Education*, *18*(3), 65-77.

Jiang, Y., Wang, W., & Kang, M. M. (2019). Research on the effect of teacher-student interaction based on behavior sequence analysis. *Modern Distance Education*, *2019*(6), 53-61.

Lemke, J. L. (1983). Thematic analysis: Systems, structures, and strategies. RSSI

Recherches Sémiotiques/Semiotic Inquiry, *3*(1-2-3-4), 159-187.

Li, C. L. (2014). *Research on classroom questioning behavior in high-quality high school geography lessons* (Master’s thesis). Inner Mongolia Normal University.

Li, S., Huang, J. J., & Liu, S. Z. (2022). Analysis of teacher-student dialogue interaction patterns and characteristics in live-streaming teaching. *Modern Distance Education Research*, *34*(4), 91-103+112.

Liu, F., & Wen, W. (2015). New trends in the PISA 2015 scientific literacy assessment framework and its implications for science education in China. *Foreign Education Research*, *42*(10), 117-128.

Lyle, S. (2008). Dialogic teaching: Discussing theoretical contexts and reviewing evidence from classroom practice. *Language and Education*, *22*(3), 222-240.

Shi, J. M. (2022). Key difficulties in implementing the new compulsory education science curriculum standards: An interview with Hu Weiping, head of the revision team of the compulsory education science curriculum standards. *People's Education*, *2022*(Z2), 38-40.

Viiri, J., & Saari, H. (2006). Teacher talk patterns in science lessons: Use in teacher education. *Journal of Science Teacher Education*, *17*(4), 347-365.

Wang, G., & Yang, F. (2024). Analysis of scientific classroom discourse from the perspective of Susan Haack's foundherentism. *Journal of Comparative Education*, *2024*(4), 148-164.

Wei, S. P. (2013). Learning analytics: Unlocking the value of educational data in the era of big data. *Modern Educational Technology*, *23*(2), 5-11.

Zhang, H., Cui, Y. L., & Yu, L. Y., et al. (2020). Research on the event-based knowledge graph of smart classroom teaching behavior based on data mining. *Distance Education Journal*, *38*(2), 80-88.

Zhong, Q. Q. (2006). Knowledge construction and teaching innovation: Social constructivist epistemology and its implications. *Global Education Outlook*, *35*(8), 12-18.

Zhou, P. H., Zhang, Y., & Yang, Q. R., et al. (2018). Analysis of classroom discourse in collaborative knowledge construction in smart classrooms—A case study of primary school science courses. *E-Education Research*, *39*(1), 20-28.

Zhu, K., Wu, Y. X., & Hu, J. Y., et al. (2024). Interdisciplinary thematic learning for knowledge construction: Model development and practical research. *Modern Educational Technology*, *34*(10), 65-74.

N. Law, G. Wong, E. Woo, & G. Jiang (Eds.) (2025). *Conference Proceedings of CITERS 2025.* Hong Kong: The University of Hong Kong.

The Impact of ChatGPT on Problem-Based Learning in Medical Education: A Comparative Study on an Anaphylaxis Tutorial

Ching Hei So[1*], Betty Yee Man Au-Yeung[1], Chris Sheung Chit Ng[1], Derrick H M Chan[1], Dickson Wing San Yiu[1], Hoi Chak Ng[1], Tsi Lok Ho[1], Zak Song[1], Shilpa Purdal[2], Kendrick Shih[3], and Philip H Li[4]
[1]LKS Faculty of Medicine, The University of Hong Kong, Hong Kong SAR
[2]School of Public Health, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong SAR
[3]Department of Ophthalmology, School of Clinical Medicine, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong SAR
[4]Department of Medicine, School of Clinical Medicine, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong SAR
[*] as895383@connect.hku.hk

Abstract: Generative AI in medical education remains promising but its effects on learner performance are unclear. Our pilot randomised controlled trial studied the impact of ChatGPT-assisted learning on anaphylaxis in undergraduate medical students. The intervention group outperformed the control in learning outcomes, highlighting potential AI dependency concerns. While ChatGPT aided knowledge retrieval, its efficacy in complex tasks was limited. Further research is crucial to understand AI-human interactions and mitigate biases.
Keywords: Anaphylaxis, generative AI, medical education, problem based learning

1. INTRODUCTION

Generative AI, particularly ChatGPT, has shown promising applications in healthcare, including passing the USMLE exams with high concordance in late 2022 [1], and later achieving over 86% accuracy with GPT-4 despite minimal prompting [2], and surpassing 75% in UKMLA [3](1,2). In clinical vignettes, ChatGPT reached above 70% accuracy on average and nearly 77% accuracy in diagnosis [4]. In addition, AI models may assist in medical note- taking, consultations, and evaluation [5]. Start-up companies such as Nabla and Abridge have also developed AI co-pilots for automated generation of medical documentation [6,7]. Concepts on general models in medical AI were recently proposed amid the recent advances of generalist large language models, where such models could potentially allow flexible human-computer interactions for dynamic task completions through self-supervised learning and training with medical domain knowledge [8]. (it has also been proposed that these language models could enable flexible human-computer interactions and dynamic task completion through self-supervised learning and training with medical knowledge)

The integration of generative AI into healthcare extends beyond clinical practice into medical education, where it holds potentials to serve as tutors, coaches, mentors, teammates, students, tools, and simulators to learners [9]. It can generate immersive patient scenarios for students [10], provide learning insights, assist in deductive reasoning [1], and generate exam-style multiple-choice questions) [11]. Commercially, GPT-4 was adopted by Khan Academy's Khanmingo chatbot which fosters critical thinking by guiding students without giving direct answers [12]. Despite the promise and growing adoption of generative AI, there is limited data on its impact on learner performance and experience.

This study aims to evaluate the impact of generative AI-assisted learning in problem-based learning (PBL) tutorials for undergraduate medical students using ChatGPT in a case of anaphylaxis as a pilot randomised control trial. We also aim to evaluate the difference between GPT-4.0 vs GPT-3.5 in solving the PBL case. PBL is widely adopted in medical education, where students develop critical thinking, self-directed learning, and practical skills by solving clinical problems in small, tutor-guided groups [13]. Given the integral role of PBL and case-based learning in medical curricula and the potential of generative AI to facilitate clinical problem solving [14], understanding the role of generative AI in PBL tutorials and its impact on students' learning efficacy and experience is crucial.

2. ACTIVITY

We explored the use of ChatGPT on a 1.5-hour PBL tutorial on Wheat-dependent exercise-induced anaphylaxis (WDEIA) adopted from the fourth year Bachelor of Medicine and Bachelor of Surgery (MBBS) curriculum at The University of Hong Kong (HKU). This case involves complex pathophysiology and requires students to integrate basic science and clinical skills to manage a rare disease they likely have not encountered before.

Nine medical students were recruited and randomly assigned into intervention (using ChatGPT, n=5) and control groups (not using ChatGPT, n=4), with each group led by a randomly assigned clinical academic tutors who are blinded with the group assignments. Students went through the PBL case with tutor's guidance as usual. At the end of the session, both tutors and students filled in a feedback form evaluating the number of fulfilled learning outcomes and their perspectives towards ChatGPT.

ChatGPT, utilising both GPT-3.5 and GPT-4 models, was instructed to solve the PBL case with minimal prompting, and the outputs were evaluated by a blinded clinical academic. The performances of GPT-3.5 and GPT-4 models were then compared with that of students. (Details of the surveys are included in Supplementary Information 1, Appendix 1-4)

3. RESULTS AND DISCUSSION

The study involved second to fourth- year MBBS students at HKU, with a predominantly male demographic (89%). A significant proportion of students had experience using generative AI (89%), with 56% having experience in prompt engineering, and 22% in coding generative AI models. In the intervention group (Group A, using ChatGPT) (n=5), 80% used the GPT-3.5 model, and 20% used both GPT-3.5 and GPT-4 models during the tutorial.

3.1 Learning Outcomes and Student Performance

The intervention group performed better in achieving learning outcomes across all domains compared to the control group (Fig 1a, Table 1). The intervention group met all learning outcomes more favourably, compared to 23% in the control group. Such differences were independent of the students' levels of participation, communication, critical thinking, and group skills, which were comparable between the groups (Fig 1b). The enhanced ability of students to achieve learning outcomes with the adoption of ChatGPT could suggest an enhanced ability to tackle clinical problems with the human-AI combination, which was also noted by Bienefeld and colleagues in managing cases in the intensive care unit setting [15].

However, it remains questionable whether this improved achievement in learning outcomes in human-AI combination could translate into better development of clinical reasoning skills. Both tutor and students in the intervention group observed that students may bypass "information synthesis" and "fail to experience the struggles of analysing the case" necessary to "develop the necessary clinical reasoning skills" (Fig 1g), which are the key components of PBL. Similar observations were reported by Hamid and colleagues in a similar study on a pharmacy PBL case, where instructors observed that easy access to prompts and responses from ChatGPT might lead to dependency and weaken the creative and critical thinking capabilities of students [16].

Such observation perhaps highlights a key concern of over-reliance when adopting ChatGPT in medical education [17]. Ideally, generative AI tools should be used critically, as the models can provide false responses, known as "hallucinations" [5,18,19]. Responses from the limitations of generative AI could lead to automation bias [21], which could compromise patient care when making clinical decisions [20,21]. This may be circumvented with improved AI education for medical students [22] or adopting bias mitigation strategies to encourage critical reflection on AI output [23]. Perhaps when students better understand the limitations of ChatGPT, they could benefit by incorporating ChatGPT's ideas into their clinical reasoning more critically.

3.2 Student perception of the PBL tutorial and ChatGPT

Both groups generally agreed that the PBL tutorial enhanced their learning, confidence, and interest in the topic, and were satisfied with the tutorial (Fig 1c). There was more variation in the intervention group's feedback regarding ChatGPT, although the overall perceptions were positive (Fig 1d). Interestingly, the control group unanimously perceived that ChatGPT may improve confidence and facilitate a smoother PBL session but not the intervention group, which could reflect an unmet optimism towards ChatGPT in PBL settings. generative AI are also not tractable due to their "black box" nature [20]. Failure to acknowledge

In both groups, most agreed that ChatGPT could assist in conceptual and knowledge-based tasks like obtaining factual knowledge, explaining and clarifying concepts, providing a second opinion, and proposing investigation & management plans (87% agreed), offering more perspectives on the case, generating differential diagnoses (78% agreed), and generating questions for knowledge checking (67% agreed) (Figure 1e). In contrast, ChatGPT was considered less helpful in more complex tasks involving more human elements and interactions including role play (only 22% agreed) and to serve as the PBL tutor (33% agreed). While the execution of complex acting tasks could indeed be more challenging due to hallucinations, recent studies showed that ChatGPT could exhibit behaviours similar to humans [24], if not more altruistic, cooperative [24], and demonstrate more empathy [25]. Interestingly, ChatGPT was perceived to be more helpful in obtaining factual knowledge by the intervention group than the control group ($p = 0.01$), which could reflect an underestimated capacity for students to retrieve factual knowledge retrieval if they are not usingChatGPT in PBL. The high agreement of ChatGPT's capacity for factual knowledge retrieval was also reported by Hamid and colleagues, where all students agreed that ChatGPT helped answer helpful in answering questions and improve understanding of the PBL tasks [16].

Overall, 67% of students favoured the incorporation of ChatGPT in this PBL tutorial, lower than what was reported by Hamid and colleagues where all students agreed that ChatGPT or similar AI should be incorporated in future PBL session [16]. This could reflect more

scepticism towards ChatGPT in our cohort, where concerns including the hindrance of "critical medical thinking", complacency, "inability to replace human thinking" were repeatedly noted by students and tutors (Fig 1g, full list of comments in Supplementary Information 1,
Appendix 5). Perhaps the question of whether or how generative AI should be incorporated in PBL and in medical education is beyond trivial. Despite its benefits observed, more work is needed to address the drawbacks of its adoption on student learning, particularly the development of critical thinking.

3.3 Comparing GPT-3.5 vs GPT-4 models

The PBL case was then fed into ChatGPT with minimal prompting to compare the responses using either GPT-3.5 or GPT-4 models. Both GPT-3.5 and GPT-4 models achieved all learning outcomes favourably, and GPT-4 achieved more favourably than GPT-3.5, consistent with other comparative studies [26,27]. (Figure 1f). The case marker could successfully match the responses to the model used while being blinded, which demonstrates noticeable differences between the models. The outstanding performances, even in the GPT-3.5 model, could reflect its capacity as a clinical case-support tools with reasonable reliability. Nevertheless, its seemingly reliable capacity could promote complacency and discourage critical thinking as observed and should therefore be used with caution.

3.4 Limitations

This study would benefit from larger sample sizes and more equal gender distribution. Inter-student and inter-tutor variability, despite adjustment via randomisation, could contribute to variations in the assessed outcomes. The learning outcomes could also be further compared by administrating a post-tutorial quiz where both control and intervention groups without LLM assistance. Besides, face-to-face sessions might yield different observations compared to the online format used in this study. While PBL tutorials could recapitulate important features of clinical problem solving, further studies are needed to evaluate the adoption of generative AI in the clinical education and practice setting.

4. CONCLUSIONS & FUTURE WORK

In conclusion, the use of ChatGPT in PBL improved students' ability to achieve learning outcomes and was perceived helpful in conceptual and knowledge-based tasks yet demonstrated concerns in over-reliance. While generative AI could support students' learning in PBL, further studies are needed to understand the AI-human interactions and to mitigate automation bias prior to its adoption.

5. FIGURES AND TABLES

Figure 1.

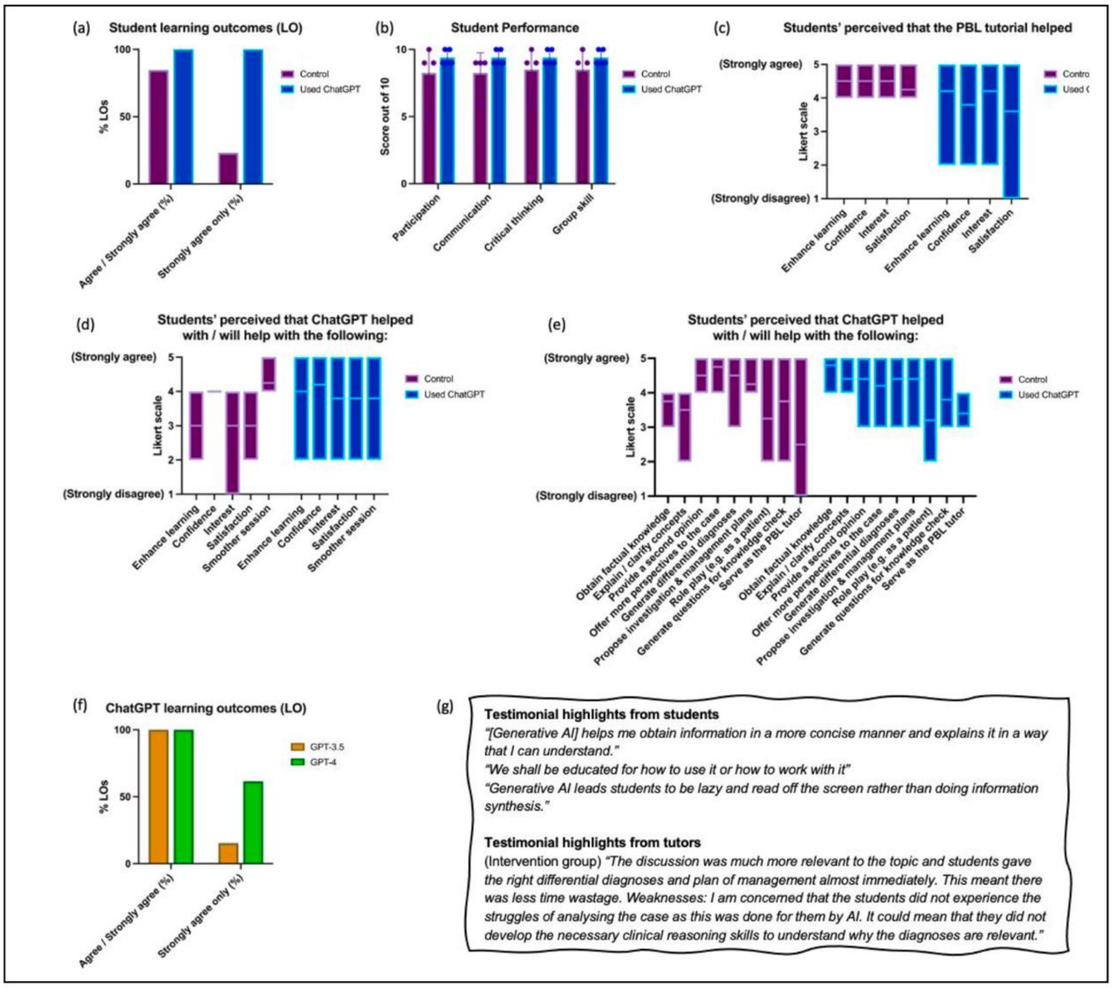


Figure 1. (a) The percentage of learning outcomes students achieved as assessed by tutors. (b) Student performances in participation, communication, critical thinking, and group skill assessed by tutors. Mean ± standard deviation are shown on the graph. (c) Student's perception of how well the PBL tutorial enhanced their learning, built their confidence, and interest towards the topic, and the satisfaction towards the tutorial. (d) Student's perception of how well the ChatGPT enhanced (for intervention group) or may enhance (for control group) their learning during the PBL session. (e) Student's perception of the roles ChatGPT played (for intervention group) or may play (for control group) in PBL sessions. (f) The percentage of learning outcomes
ChatGPT achieved (GPT-3.5 or GPT-4) as assessed by tutor. (c-f) The range (minimum to maximum) and the mean were shown on the graphs. (g) Testimonial highlights from students and tutors received at the end of the PBL session.

Table 1.

	Control	Intervention	GPT-3.5	GPT-4
I. Human Biology in Health and Disease				
1. Identify causes of syncope in adult patients.	5	5	4	5
2. Describe the pathophysiology of type I hypersensitivity reactions and anaphylaxis.	4	5	4	5

3. Describe the clinical manifestations of type I hypersensitivity reactions and anaphylaxis.	4	5	4	4
4. Understand the diagnostic approach and management of exercise-induced anaphylaxis.	4	5	4	4
5. Understand the management of anaphylaxis.	4	5	4	4
II. Professional Skills: Diagnostic, Problem Solving, Effective Communication and Clinical Management				
1. Describe the evaluation of a patient with syncope.	4	5	4	4
2. Understand the different causes of urticaria.	4	5	5	4
3. Recognize the importance of a comprehensive medical history in assessing patients and exclusion of more common diagnoses before diagnosing uncommon diseases.	5	5	4	5
4. Appreciate the role of skin tests in the allergy assessment and its interpretation.	4	5	4	5
III. Population Health, Health Services, Economics and Policy				
1. Understand the disease burden of allergy and anaphylaxis and its impact on individuals' quality of life.	5	5	4	5
2. Understand the importance of public awareness and policies in managing food allergy.	3	5	4	5
IV. Medical Ethics, Professional Attitudes and Behaviour				
1. Understand the importance of patient communication and education in explaining diagnosis and management.	4	5	4	5
2. Understand the importance of multidisciplinary collaboration when managing patients with allergy.	3	5	5	5

Table 1. Table showing the ability of the control group, intervention group, GPT-3.5, and GPT-4 in achieving the learning outcomes as assessed by tutors or case marker. 5 = Strongly agree, 4 = Agree, 3 = Neutral, 2 = Disagree, 1 = Strongly disagree.

ACKNOWLEDGEMENT
We would like to acknowledge the support from Innovation Academy towards this project. This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the Ethics Committee of the University. Informed consent was obtained from all individual participants included in the study.

SUPPLEMENTARY INFORMATION

1. Appendix 1 – Student evaluation form

The following outlines the questions included in the tutor evaluation form completed after the PBL session.

Part I – Basic information (For both intervention & control group)
1. Gender (Options: Female / Male / Others / Prefer not to say)
2. Cohort (2023-24 academic year) (Options: MBBS 26’ / MBBS 27’ / MBBS 28’; represents the graduation year)

Part II – Tutorial experience (For intervention group only)
1. How did the PBL tutorial impact your learning experience of the case? (Options: Strongly disagree / Disagree / Neutral / Agree / Strongly agree)
a. The tutorial enhanced my learning of this case.
b. I felt confident in tackling the case after the tutorial.
c. I am more interested in the case because of the tutorial
d. I am satisfied with the use of PBL tutorial in learning this case.
2. Which large language model (LLM) did you use? (Select all that apply) (Options: Strongly disagree / Disagree / Neutral / Agree / Strongly agree)
a. ChatGPT (GPT-3.5)
b. ChatGPT (GPT-4)
c. Not applicable

d. Other…
3. How did the use of ChatGPT impact your learning experience during the PBL case? (Options: Strongly disagree / Disagree / Neutral / Agree / Strongly agree)
a. ChatGPT enhances my learning in this case.
b. I felt confident in tackling the case using ChatGPT.
c. I am more interested in the case because of ChatGPT
d. I am satisfied with the use of ChatGPT in this case.
e. ChatGPT made the PBL session smoother
4. To what extent did ChatGPT help you achieve the following learning objectives? (Options: Strongly disagree / Disagree / Neutral / Agree / Strongly agree)
a. Identify causes of syncope in adult patients.
b. Describe the pathophysiology of type I hypersensitivity reactions and anaphylaxis.
c. Describe the clinical manifestations of type I hypersensitivity reactions and anaphylaxis.
d. Understand the diagnostic approach and management of exercise-induced anaphylaxis.
e. Understand the management of anaphylaxis.
f. Describe the evaluation of a patient with syncope
g. Understand the different causes of urticaria.
h. Recognize the importance of a comprehensive medical history in assessing patients and exclusion of more common diagnoses before diagnosing uncommon diseases.
i. Appreciate the role of skin tests in the allergy assessment and its interpretation.
j. Understand the disease burden of allergy and anaphylaxis and its impact on individuals' quality of life.
k. Understand the importance of public awareness and policies in managing food allergy.
l. Understand the importance of patient communication and education in explaining diagnosis and management.
m. Understand the importance of multidisciplinary collaboration when managing patients with allergy.
5. In what ways can ChatGPT facilitate your learning in the tutorial? (Options: Strongly disagree / Disagree / Neutral / Agree / Strongly agree)
a. Obtain factual knowledge
b. Explain / clarify concepts
c. Provide a second opinion
d. Offer more perspectives to the case
e. Generate differential diagnoses
f. Propose investigation and management plans
g. Role play (e.g. as a patient)
h. Generate questions for knowledge check
i. Serve as the PBL tutor
6. Are there other ways ChatGPT can facilitate your learning? (Open response)
7. To what extent do you agree to the following statements? (Options: Strongly disagree / Disagree / Neutral / Agree / Strongly agree)
a. I participate and engage in the tutorial discussion
b. I would prefer the incorporation of ChatGPT in this tutorial.
Part II – Tutorial experience (For control group only)
1. How was your learning experience in the PBL tutorial? (Options: Strongly disagree / Disagree / Neutral / Agree / Strongly agree)
a. The tutorial enhanced my learning of this case.
b. I felt confident in tackling the case after the tutorial.

c. I am more interested in the case because of the tutorial
d. I am satisfied with the use of PBL tutorial in learning this case.
2. H How do you think the use of ChatGPT may impact your learning experience for the PBL case? (Options: Strongly disagree / Disagree / Neutral / Agree / Strongly agree)
a. ChatGPT enhances my learning in this case.
b. I felt confident in tackling the case using ChatGPT.
c. I am more interested in the case because of ChatGPT
d. I am satisfied with the use of ChatGPT in this case.
e. ChatGPT made the PBL session smoother
3. In what ways can ChatGPT facilitate your learning in the tutorial? (Options: Strongly disagree / Disagree / Neutral / Agree / Strongly agree)
a. Obtain factual knowledge
b. Explain / clarify concepts
c. Provide a second opinion
d. Offer more perspectives to the case
e. Generate differential diagnoses
f. Propose investigation and management plans
g. Role play (e.g. as a patient)
h. Generate questions for knowledge check
i. Serve as the PBL tutor
4. Are there other ways ChatGPT can facilitate your learning? (Open response)
5. To what extent do you agree to the following statements? (Options: Strongly disagree / Disagree / Neutral / Agree / Strongly agree)
a. I participate and engage in the tutorial discussion
b. I would prefer the incorporation of ChatGPT in this tutorial.
Part III – For both intervention & control group
1. How much experience do you have in using generative AI / LLMs? (Options: Strongly disagree / Disagree / Neutral / Agree / Strongly agree)
a. I have experience in using generative AI / LLM
b. I have experience in prompt engineering to obtain productive response from AI.
c. I have experience in coding generative AI programmes
2. To what extent do you agree to the following statements? (Options: Strongly disagree / Disagree / Neutral / Agree / Strongly agree)
a. Generative AI helps me obtain more relevant information from AI than other means (e.g. from Google search / 'Goddisk notes').
b. I have concerns with the use of generative AI in medical education.
c. Students should have access to AI during PBL sessions.
d. Students should have access to AI during assessments.
e. More training should be provided to staff & students on the use of generative AI.
f. Generative AI does more good than harm in medical education.
3. Please write down other comments on the session / any thoughts on the use of generative AI in medical education below. (Optional) (Open response)
4. Moving to a broader perspective, how do you think generative AI can improve medical education? If you think generative AI is beneficial in medical education, how can the curriculum better incorporate the use of generative AI? (Optional) (Open response)

1. Appendix 2 – Tutor evaluation form

The following outlines the questions included in the tutor evaluation form completed after the PBL session.

Part I – Student evaluation
Please kindly evaluate the student's performance in the mock PBL session below. Your ratings will be used for quantitative analysis of the variations in student performance between the control and intervention groups.
1. Student A-D/E (Each student is ranked according to the following criteria) (Options: 1 (Poor) / 2 (Poor) / 3 (Poor) / 4 (Unsatisfactory) / 5 (Unsatisfactory) / 6 (Average) / 7 (Above average) / 8 (Above average) / 9 (Excellence) / 10 (Excellence))
a. Participation
b. Communication
c. Critical thinking
d. Group skill
2. Please indicate the extent to which you agree students as a group achieved the learning outcomes:
a. Identify causes of syncope in adult patients.
b. Describe the pathophysiology of type I hypersensitivity reactions and anaphylaxis.
c. Describe the clinical manifestations of type I hypersensitivity reactions and anaphylaxis.
d. Understand the diagnostic approach and management of exercise-induced anaphylaxis.
e. Understand the management of anaphylaxis.
f. Describe the evaluation of a patient with syncope
g. Understand the different causes of urticaria.
h. Recognize the importance of a comprehensive medical history in assessing patients and exclusion of more common diagnoses before diagnosing uncommon diseases.
i. Appreciate the role of skin tests in the allergy assessment and its interpretation.
j. Understand the disease burden of allergy and anaphylaxis and its impact on individuals' quality of life.
k. Understand the importance of public awareness and policies in managing food allergy.
l. Understand the importance of patient communication and education in explaining diagnosis and management.
m. Understand the importance of multidisciplinary collaboration when managing patients with allergy.
Part II – General questions for reflection
1. How difficult was it for you to conduct the PBL, compared to others you've conducted? (Scale bar: 1 (Much more difficult) … 3 (No perceivable change) … 5 (Much easier))
2. How comfortable did students seem in navigating the complexities of the case during the PBL, compared to others you've conducted? (Scale bar: 1 (Much less comfortable) … 3 (No perceivable change) … 5 (Much more comfortable))
3. How would you rate the overall flow/smoothness of the PBL, compared to others you've conducted? (Scale bar: 1 (Much less smooth and disjointed) … 3 (No perceivable change) … 5 (Much smoother and more coherent))
4. How engaged/actively participating are the students during the PBL, compared to others you've conducted? (Scale bar: 1 (Much less engaged) … 3 (No perceivable change) … 5 (Much more engaged))
5. How would you describe the pace of the PBL session in terms of covering the desired content? (Scale bar: 1 (Too fast) … 3 (Just right) … 5 (Too slow))
6. How effectively did students collaborate and work as a team during the PBL, compared to others you've conducted? (Scale bar: 1 (Much poorer collaboration / synergy) … 3 (No perceivable change) … 5 (Much better collaboration / synergy))
7. How would you rate the level of independent thinking/originality of ideas the students

exhibited during the sessions, compared to others you've conducted? (Scale bar: 1 (Much more dependent on external information) … 3 (No perceivable change) … 5 (Much more independent / original))

8. Did the student show evidence of genuine understanding of the topics/concepts presented during the PBL, compared to others you've conducted? For example, consider whether the student was just reading off a text from the internet or showing signs that they actually have understood and digested the content they were presenting. (Scale bar: 1 (Much poorer evidence of genuine understanding) … 3 (No perceivable change) … 5 (Much better evidence of genuine understanding))

9. How would you rate the overall student performance during the PBL, compared to others you've conducted? (Scale bar: 1 (Much poorer) … 3 (No perceivable change) … 5 (Much better))

10. Were there any unique strengths or weaknesses of this PBL session compared to others you've facilitated? (Open response)

11. Do you have any other comments on the PBL you conducted/observed?

3. Appendix 3 – ChatGPT evaluation form

The following outlines the questions and the options included in the evaluation form for the response generated by ChatGPT.

1. Please indicate the extent to which you agree Response A achieved the learning outcomes (Options: Strongly disagree / Disagree / Neutral / Agree / Strongly agree)

a. Identify causes of syncope in adult patients.
b. Describe the pathophysiology of type I hypersensitivity reactions and anaphylaxis.
c. Describe the clinical manifestations of type I hypersensitivity reactions and anaphylaxis.
d. Understand the diagnostic approach and management of exercise-induced anaphylaxis.
e. Understand the management of anaphylaxis.
f. Describe the evaluation of a patient with syncope
g. Understand the different causes of urticaria.
h. Recognize the importance of a comprehensive medical history in assessing patients and exclusion of more common diagnoses before diagnosing uncommon diseases.
i. Appreciate the role of skin tests in the allergy assessment and its interpretation.
j. Understand the disease burden of allergy and anaphylaxis and its impact on individuals' quality of life.
k. Understand the importance of public awareness and policies in managing food allergy.
l. Understand the importance of patient communication and education in explaining diagnosis and management.
m. Understand the importance of multidisciplinary collaboration when managing patients with allergy.

2. Please indicate the extent to which you agree Response B achieved the learning outcomes (Options: Strongly disagree / Disagree / Neutral / Agree / Strongly agree)

a. Identify causes of syncope in adult patients.
b. Describe the pathophysiology of type I hypersensitivity reactions and anaphylaxis.
c. Describe the clinical manifestations of type I hypersensitivity reactions and anaphylaxis.
d. Understand the diagnostic approach and management of exercise-induced anaphylaxis.
e. Understand the management of anaphylaxis.
f. Describe the evaluation of a patient with syncope.
g. Understand the different causes of urticaria.

h. Recognize the importance of a comprehensive medical history in assessing patients and exclusion of more common diagnoses before diagnosing uncommon diseases.
i. Appreciate the role of skin tests in the allergy assessment and its interpretation.
j. Understand the disease burden of allergy and anaphylaxis and its impact on individuals' quality of life.
k. Understand the importance of public awareness and policies in managing food allergy.
l. Understand the importance of patient communication and education in explaining diagnosis and management.
m. Understand the importance of multidisciplinary collaboration when managing patients with allergy.
3. The two responses were generated by either GPT-3.5 or GPT-4. Which of the following ChatGPT models do you think was used to generate Response A and Response B? (Options: GPT-3.5, GPT-4)
a. Response A
b. Response B
4. Do you have any comments regarding the responses or the case?

4. Appendix 4 – Prompting of ChatGPT generated response using either GPT-3.5 or GPT-4

The following outlines the prompt used for ChatGPT generated responses using GPT-3.5 or GPT-4.

[Prompt 1 begins]
This is a problem-based learning exercise for medical students. We would like your response as you go through the case. Here are the guiding questions:
I. Human Biology in Health and Disease
Identify causes of syncope in adult patients.
Describe the pathophysiology of type I hypersensitivity reactions and anaphylaxis.
Describe the clinical manifestations of type I hypersensitivity reactions and anaphylaxis.
Understand the diagnostic approach and management of exercise-induced anaphylaxis.
Understand the management of anaphylaxis.
II. Professional Skills: Diagnostic, Problem Solving, Effective Communication and Clinical Management
Describe the evaluation of a patient with syncope.
Understand the different causes of urticaria.
Recognize the importance of a comprehensive medical history in assessing patients and exclusion of more common diagnoses before diagnosing uncommon diseases.
Appreciate the role of skin tests in the allergy assessment and its interpretation.
III. Population Health, Health Services, Economics and Policy Understand the disease burden of allergy and anaphylaxis and its impact on individuals' quality of life.
Understand the importance of public awareness and policies in managing food allergy.
IV. Medical Ethics, Professional Attitudes and Behaviour
Understand the importance of patient communication and education in explaining diagnosis and management.
Understand the importance of multidisciplinary collaboration when managing patients with allergy.
At the end of each page, provide reflection on each case with respect to the learning objectives. I will be providing the case.
[Prompt 1 ends]

After this prompt, each section of the PBL case is fed into ChatGPT until a response is generated for all sections. The PBL case and the ChatGPT generated response can be provided by the authors upon reasonable request.

5. Appendix 5 – List of open response comments from the student & tutor evaluation forms

Part A – Open responses from the student evaluation form

1. Please write down other comments on the session / any thoughts on the use of generative AI in medical education below.

- Dependence on AI restricts development of critical medical thinking
- I think that generative AI leads students to be lazy and read off the screen rather than doing information synthesis. Also, there can sometimes be inaccurate information that may be quoted during synthesis of information, which medical students may not catch and thus accidentally believe. However, it can be useful to summarise large chunks of information that can be difficult to digest. It can also provide more aspects of information from different sources when compared to a simple Google search. I think that while it should not be involved in PBL tutorials and examinations, it is still a hugely useful asset to have while revising and formulating notes, as well as for further learning (when combined with other reputable tools e.g. UpToDate).
- I think it good as a powerful search engine but cannot replace human thinking in current stage. It is effective and accurate in gathering information but not for systematic interpretation. However, we shall be educated for how to use it or how to work with it as it is continually evolving and may be much more intelligent when I become a doctor.

2. Moving to a broader perspective, how do you think generative AI can improve medical education? If you think generative AI is beneficial in medical education, how can the curriculum better incorporate the use of generative AI?

- Helps me obtain information in a more concise manner and explains it in a way that I can understand. I do not consider the use of AI to be harmful as I think it genuinely helps me understand things better than if I were to not use AI.
- To help with the organization of notes and ideas. Production of flash cards from my lecture notes. Integrating into PBL May harm the thinking process and knowledge gathering skills in our learnings. But in future practice it definitely helps to provide another perspective and as a check for negligence.

Part B – Open responses from the tutor evaluation form

1. Were there any unique strengths or weaknesses of this PBL session compared to others you've facilitated?

- (Control group tutor) The complexity of the case was much higher than the usual cases, time constraints were there. Students were engaging and enthusiastic overall.
- (Intervention group tutor) Strengths: The discussion was much more relevant to the topic and students gave the right differential diagnoses and plan of management almost immediately. This meant there was less time wastage. Weaknesses: I am concerned that the students did not experience the struggles of analysing the case as this was done for them by AI. It could mean that they did not develop the necessary clinical reasoning skills to understand why the diagnoses are relevant. It was an ultrafast and smooth discussion, but perhaps PBLs should be more chaotic and require students to struggle in order to grow and develop as clinicians.

2. Do you have any other comments on the PBL you conducted/observed?

- (Control group tutor) Students were engaging and enthusiastic overall

 (Intervention group tutor) It was an ultrafast and smooth discussion, but perhaps PBLs should be more chaotic and require students to struggle in order to grow and develop as clinicians.

REFERENCES

Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, et al. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. PLOS Digital Health. 2023;2(2):e0000198.

Nori H, King N, McKinney SM, Carignan D, Horvitz E. Capabilities of GPT-4 on Medical Challenge Problems. arXiv. 2023;2303.13375.

Lai UH, Wu KS, Hsu TY, Kan JKC. Evaluating the performance of ChatGPT-4 on the United Kingdom Medical Licensing Assessment. Front Med (Lausanne). 2023;10:1240915.

Rao A, Pang M, Kim J, Kamineni M, Lie W, Prasad AK, et al. Assessing the Utility of ChatGPT Throughout the Entire Clinical Workflow: Development and Usability Study. J Med Internet Res. 2023;25:e48659.

Lee P, Bubeck S, Petro J. Benefits, Limits, and Risks of GPT-4 as an AI Chatbot for Medicine. New England Journal of Medicine. 2023;388(13):1233-9.

Nabla. Nabla Copilot 2023 [2023 July 3]. Available from: https://www.nabla.com.

Abridge AI Inc. Abridge 2023 [2023 July 3]. Available from: https://www.abridge.com.

Moor M, Banerjee O, Abad ZSH, Krumholz HM, Leskovec J, Topol EJ, et al. Foundation models for generalist medical artificial intelligence. Nature. 2023;616(7956):259-65.

Mollick ER, Mollick L. Assigning AI: Seven Approaches for Students, with Prompts. SSRN. 2023;ssrn.4475995.

Karabacak M, Ozkara BB, Margetis K, Wintermark M, Bisdas S. The Advent of Generative Language Models in Medical Education. JMIR Med Educ. 2023;9:e48163.

Cheung BHH, Lau GKK, Wong GTC, Lee EYP, Kulkarni D, Seow CS, et al. ChatGPT versus human in generating medical graduate exam multiple choice questions-A multinational prospective study (Hong Kong S.A.R., Singapore, Ireland, and the United Kingdom). PLoS One. 2023;18(8):e0290691.

Khan Academy. Khanmigo Education AI Guide 2023 [2023 July 3]. Available from: https://www.khanacademy.org/khan-labs.

Trullàs JC, Blay C, Sarri E, Pujol R. Effectiveness of problem-based learning methodology in undergraduate medical education: a scoping review. BMC Medical Education. 2022;22(1):104.

Divito CB, Katchikian BM, Gruenwald JE, Burgoon JM. The tools of the future are the challenges of today: The use of ChatGPT in problem-based learning medical education. Medical Teacher. 2024;46(3):320-2.

Bienefeld N, Keller E, Grote G. Human-AI Teaming in Critical Care: A Comparative Analysis of Data Scientists' and Clinicians' Perspectives on AI Augmentation and Automation. J Med Internet Res. 2024;26:e50130.

Hamid H, Zulkifli K, Naimat F, Che Yaacob NL, Ng KW. Exploratory study on student perception on the use of chat AI in process-driven problem-based learning. Currents in Pharmacy Teaching and Learning. 2023;15(12):1017-25.

Boscardin CK, Gin B, Golde PB, Hauer KE. ChatGPT and Generative Artificial Intelligence for Medical Education: Potential Impact and Opportunity. Academic Medicine. 2024;99(1).

Dash D, Thapa R, Banda JM, Swaminathan A, Cheatham M, Kashyap M, et al.

Evaluation of GPT-3.5 and GPT-4 for supporting real-world information needs in healthcare delivery. arXiv. 2023;2304.13714.

Au Yeung J, Kraljevic Z, Luintel A, Balston A, Idowu E, Dobson RJ, et al. AI chatbots not yet ready for clinical use. Front Digit Health. 2023;5:1161098.

Khera R, Simon MA, Ross JS. Automation Bias and Assistive AI: Risk of Harm From AI-Driven Clinical Decision Support. JAMA. 2023;330(23):2255-7.

Abdelwanis M, Alarafati HK, Tammam MMS, Simsekler MCE. Exploring the risks of automation bias in healthcare artificial intelligence applications: A Bowtie analysis. Journal of Safety Science and Resilience. 2024.

Nguyen T. ChatGPT in Medical Education: A Precursor for Automation Bias? JMIR Med Educ. 2024;10:e50174.

Kostick-Quenet KM, Gerke S. AI in the hands of imperfect users. npj Digital Medicine. 2022;5(1):197.

Mei Q, Xie Y, Yuan W, Jackson MO. A Turing test of whether AI chatbots are behaviorally similar to humans. Proceedings of the National Academy of Sciences.2024;121(9):e2313925121.

Ayers JW, Poliak A, Dredze M, Leas EC, Zhu Z, Kelley JB, et al. Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum. JAMA Internal Medicine. 2023;183(6):589-96.

Wang G, Gao K, Liu Q, Wu Y, Zhang K, Zhou W, et al. Potential and Limitations of ChatGPT 3.5 and 4.0 as a Source of COVID-19 Information: Comprehensive Comparative Analysis of Generative and Authoritative Information. J Med Internet Res. 2023;25:e49771.

Lee TJ, Rao AK, Campbell DJ, Radfar N, Dayal M, Khrais A. Evaluating ChatGPT-3.5 and ChatGPT-4.0 Responses on Hyperlipidemia for Patient Education. Cureus. 2024;16(5):e61067.

N. Law, G. Wong, E. Woo, & G. Jiang (Eds.) (2025). *Conference Proceedings of CITERS 2025.*
Hong Kong: The University of Hong Kong.

Cultivating Collaborative Problem-Solving Skills in the Era of Human-AI Interaction: The Role of ChatGPT

Binghao Tu[1*], Zengyi Yu
[1] College of Education, Zhejiang University of Technology, China
[*] tubinghao1103@163.com

Abstract: In the context of technological innovation driving educational technology development, collaborative problem-solving (CPS) and innovation skills are crucial for high-quality talent. This study focuses on cultivating human-AI collaborative learning abilities using ChatGPT across different knowledge contexts and learning stages. It tracks students' knowledge mastery and collaborative learning processes with ChatGPT, employing paired t-tests, independent t-tests, ANOVA, and qualitative feedback analysis to assess the impact on CPS skills. The findings show that human-AI interaction significantly enhances students' collaborative learning abilities, providing substantial support for educators in teaching.
Keywords: Collaborative Problem-Solving, Human-AI Interaction, ChatGPT

1. INTRODUCTION

In the current educational landscape, Collaborative Problem Solving (CPS) is a crucial skill for 21st-century learners and has garnered significant attention from the education community (Raiyn, J. et al.,2015). Artificial intelligence models have positively impacted the resolution of complex problems. For instance, Yu et al. (2023) developed an intelligent diagnosis model for Alzheimer's disease using neural networks, highlighting AI's potential for real-time analysis and decision support in complex scenarios (Yu, Z et.al.,2024). Applying such AI models to education can enhance the precision and timeliness of student performance monitoring and assessment, thereby supporting CPS skill development through Human-AI Collaboration (HAC). However, current teaching practices and research methods are still largely limited to traditional classroom instruction. He Yingzhao and Li Huijia integrated knowledge mapping and scholar profiling to select online resources (Guan, H., 2022), but real-time monitoring of students' knowledge mastery and tracking their progress on specific items have been overlooked. Consequently, existing methods often lack the precision and efficiency needed to effectively develop CPS competencies. Integrating ChatGPT technology with efforts to develop learners' collaborative problem-solving skills is highly important. ChatGPT, as a large language model, provides real-time feedback, supports inquiry, and simulates peer interaction. These features align well with educational goals such as improving communication, critical thinking, and knowledge co-construction.

This study addresses these deficiencies by examining the interplay between human-machine collaboration and CPS development, as well as the role of knowledge tracking in enhancing these abilities. And this study will build on the findings and theories of previous scholars and innovatively incorporate knowledge mapping to determine learner knowledge acquisition. The research is guided by two central research questions (RQs):

RQ1. What is the impact of Knowledge Graphs and individual knowledge mastery on LLM-generated text?
RQ2. What is the impact of using Knowledge Graphs and individual knowledge mastery

on students' interaction with large language models in HAC?

To address these objectives, this study employs ChatGPT technology to enhance human-machine interaction and systematically explores its potential to improve students' CPS skills. It aims to extend the methodological framework for developing CPS skills and address the limitations of traditional knowledge assessment methods. The research will produce a comprehensive report on the role of HAC in promoting CPS skills, contributing to educational theory and practice.

2. LITERATURE REVIEW

In recent years, the integration of digital and intelligent technologies has transformed the advancement of Collaborative Problem Solving (CPS) frameworks in the context of digitalization, artificial intelligence, networking, and big data. Yao et al. (2023) discussed advancements in 3D semantic understanding, which provide a foundation for enhancing real-time student monitoring through intelligent systems, dynamically assessing students' knowledge acquisition and filling gaps in traditional teaching methods (Yao, J. et.al,2023). As we enter the era of intelligent computing, the fusion of big data, artificial intelligence, cloud computing, digital twins, and other technologies is reshaping the learning environment (Barandiaran, X er al,2024). Yao et al. (2024) introduced techniques focusing on bird's-eye view object detection, which could be adapted to track student interaction with intelligent systems and knowledge graphs (Yao, J. et al.,2023). This enables high-level analysis of student performance across diverse learning contexts, allowing for more accurate and personalized learning pathways.Some reachers evaluated teacher teaching posture using AI-integrated posture recognition, illustrating how intelligent technologies can be integrated into educational environments to improve real-time monitoring of both teaching and learning behaviors (Yu, Z et al.,2023). This approach reinforces the value of human-machine collaboration in evaluating and enhancing CPS skills. However, these methods require further exploration as they are gradually being promoted.

3. METHODOLOGY

This study evaluated collaborative learning efficiency using four metrics: Task Completion Time (TCT), Interaction Density (ID), Contribution Balance Index (CBI), and Solution Quality per Time (SQ/T). These indicators collectively reflect how the integration of Knowledge Graphs (KG) and Personal Knowledge Proficiency (Pi) enhances the speed, equity, and quality of group problem solving.To formalize these effects, the interactions among Large Language Models (LLMs), KG, Pi, prompts (Pr), knowledge points (K), and generated text (Text) can be modeled through logical formulas. KG structures domain knowledge, while Pi enables adaptive support. Their combination informs LLM-generated content, improving relevance and complexity. This framework supports the design of intelligent learning systems that foster personalized, efficient, and collaborative learning.

3.1. Variables and Definitions

LLM: Large Language Model, which processes inputs and generates outputs based on acquired knowledge.

KG: Knowledge Graph, encoding the structure and relationships among knowledge points.

Pi: Personal Knowledge Proficiency, denoting an individual's level of mastery over a specific knowledge point.

Pr: Prompt, serving as the input to the LLM and encompassing context and questions related to knowledge points.

K: Specific Knowledge Points, signifying particular knowledge points within the knowledge graph.

Text: The output generated by the LLM in response to the input prompt and knowledge graph.

3.2. Formula Representation and Relationships

The formula representation and relationships can be formulated as follows:

$$\forall x(Text(x) \longleftrightarrow \exists y(LLM(y,Pr(x),k(x)) \land KG(y,Pi(x),k(x)))) \quad (1)$$

In the formula, $\forall x$ $(Text(x) \leftrightarrow \ldots)$ denotes each instance of generated text, indicating that the existence of the text is logically contingent upon the fulfillment of specific conditions. The term $\exists y(LLM(y,Pr(x),k(x)) \land KG(y,Pi(x),k(x)))$ signifies the presence of a particular large language model y that processes the prompt Pr and the knowledge point k, taking into account the interactions between the knowledge graph and personal mastery. $LLM(y,Pr(x),k(x))$ indicates that the large language model y generates text based on the prompt $Pr(x)$ and the specific knowledge point $k(x)$, reflecting the standard text generation process. $KG(y,Pi(x),k(x))$ represents the knowledge graph KG, which, in conjunction with the model y, refines the output according to the learner's mastery level Pi of the specific knowledge point k. This ensures that the generated text is tailored and contextually appropriate, aligning with the learner's individual knowledge.

Overall Relationship:

The entire system's relationship can be described as:

$$Text* = LLM(Pr, Pi, KG, k) \quad (2)$$

However, in the majority of instances, human-AI collaboration frequently lacks KG (Knowledge Graph) and Pi (Personal Knowledge Proficiency), giving rise to the subsequent formula.

$$Text = LLM(Pr, k) \quad (3)$$

Within the paradigm of human-AI collaborative learning, the interplay among Text, LLM (Large Language Model), Pr (Prompt), Pi (Personal Knowledge Proficiency), KG (Knowledge Graph), and k (Specific Knowledge Points) can be rigorously defined through a logical predicate framework. This interplay is predicated on the synergistic interaction between these components, ensuring that the generated output, Text, is not solely contingent upon the input prompt but also dynamically modulated in accordance with individualized knowledge proficiency and the structured interrelationships among knowledge points.

4. EXPERIMENT

4.1. Experiment Design

This study aims to evaluate the impact of integrating a knowledge graph (KG) and individual mastery level (Pi) into LLM-generated text on its quality and collaborative learning outcomes among junior high school students. The hypothesis (H1) posits that LLM-generated text incorporating KG and Pi will enhance student collaboration, while the null hypothesis (H0) suggests no significant difference in outcomes between text with or without KG and Pi. Sixty junior high students aged 12-15 with similar proficiency levels will participate and be randomly assigned to two groups: the control group using LLM-generated text without KG and Pi, and the experimental group using text that integrates KG and Pi. Materials include ChatGPT-4o for generating collaborative text, a knowledge graph representing domain knowledge, and Pi capturing each student's mastery of specific knowledge points.

4.2. Experimental Preparation

Students will take a pre-test to assess their baseline knowledge in a given subject. During the experimental sessions, the control group will collaborate using LLM-generated text based solely on prompt inputs, without KG or Pi. The experimental group will use LLM-generated text that incorporates KG and Pi, allowing the text to adapt to each student's knowledge proficiency and provide contextually relevant support. After collaboration, all students will take a post-test to measure knowledge acquisition and collaborative learning outcomes. Interaction quality and student engagement will be observed throughout the sessions, and feedback on the perceived utility of the LLM-generated text will be collected from both groups.

4.3. Analysis

Quantitative anasis will use paired t-tests for within-group pre- and post-test comparisons, and independent t-tests for differences between control and experimental groups. Collaboration quality, measured by interaction frequency, idea generation, and problem-solving contributions, will be analyzed using ANOVA. Qualitative feedback will assess the perceived usefulness of the LLM-generated text.

5. RESULT

This study compared the impact of LLM-generated text on student collaboration, with and without Knowledge Graph (KG) and Personal Knowledge Proficiency (Pi). Results showed the experimental group using KG and Pi significantly outperformed the control group. Specifically, they showed enhanced knowledge retention, collaborative efficiency, and mastery. The integration of KG and Pi provided clearer cognitive frameworks, facilitating complex problem-solving and innovative thinking. Additionally, structured knowledge and personalized support enhanced learning material specificity and increased student motivation and engagement, yielding more effective learning outcomes.

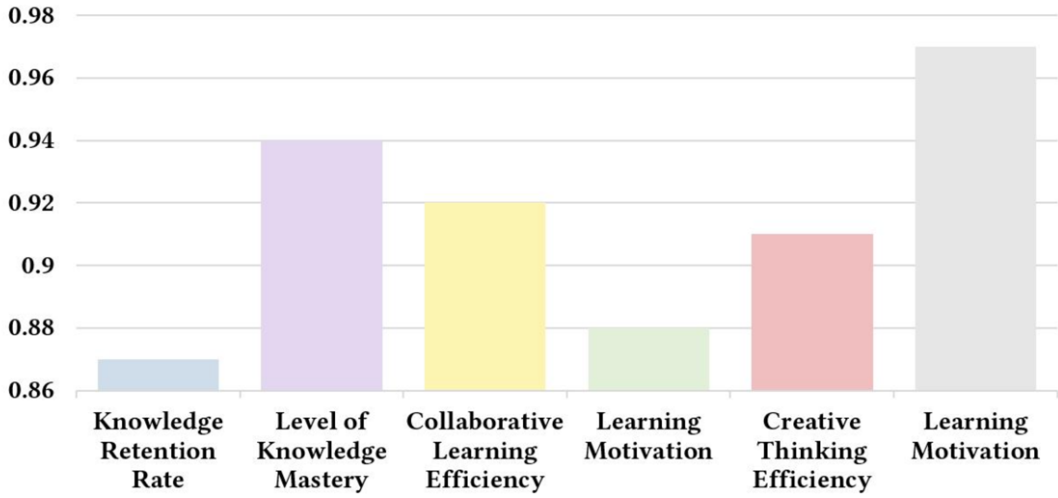


Fig 1. The Impact of Using ChatGPT-4o with Pi and KG

Based on these results, we recommend further exploration of integrating Knowledge Graph (KG) and Personal Knowledge Proficiency (Pi) into educational technology applications. This approach not only enhances students' knowledge retention and mastery but also significantly improves collaborative learning efficiency and effectiveness, promoting more efficient problem-solving and thinking skills. Additionally, personalized learning support can enhance student motivation and engagement.

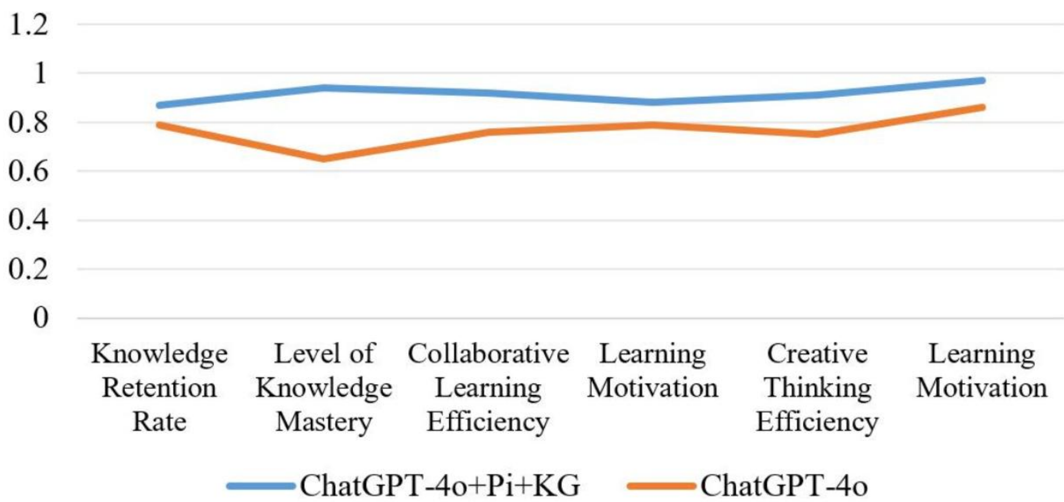


Fig 2. Comparison Between ChatGPT-4o+Pi+KG and ChatGPT-4o

As depicted in Figure 1 and Figure 2, the comparison between ChatGPT-4o (without KG and Pi) and ChatGPT-4o + Pi + KG (with KG and Pi) reveals that the experimental group (ChatGPT-4o + Pi + KG) outperformed the control group (ChatGPT-4o) in knowledge retention, demonstrating the effectiveness of integrating KG and Pi in supporting long-term information retention.

The findings of this study have significant implications for the application of AI-driven tools in educational settings. The integration of Knowledge Graphs (KG) and Personal Knowledge Proficiency (Pi) into large language models (LLMs) like ChatGPT-4o not only enhances individual knowledge retention and mastery but also optimizes collaborative learning processes. In practice, educational platforms can embed KG to structure domain-specific knowledge hierarchically, enabling learners to form clearer cognitive schemas. Simultaneously, Pi can be used to tailor instructional content to learners' existing knowledge states, thus fostering personalized scaffolding and promoting deeper engagement. By combining these two elements, educators and developers can create intelligent learning environments that support efficient problem-solving, stimulate higher-order thinking, and improve group learning outcomes. This approach can be applied in intelligent tutoring systems, collaborative learning platforms, and adaptive curricula design, ultimately advancing personalized and collaborative learning paradigms in both formal and informal educational contexts.

6. CONCLUSION

This research explores the integration of Knowledge Graphs (KG) and students' knowledge mastery with LLM-generated text. By analyzing the interactions between the LLM, KG, Personal Knowledge Proficiency (Pi), context-specific Prompts (Pr), and knowledge (K), the study shows that these integrations significantly enhance Collaborative Problem-Solving skills. Yu et al. (2024) also highlight the role of structured knowledge representations in improving system performance and personalization in GNN-based recommender systems . Their research aligns with our findings, showing how knowledge graphs can enhance the understanding of individual knowledge mastery and improve recommendation effectiveness. This supports the use of KGs in educational settings to tailor learning pathways and foster collaboration between students and AI, offering insights for future advancements in educational technology.

References

Raiyn, J., & Tilchin, O. (2015). Higher-Order Thinking Development through Adaptive Problem-based Learning. Journal of Education and Training Studies, 3(4).

Yu, Z., Wang, Y., Hu, M., & Wang, Z. (2024). Integrating Social and Knowledge Graphs in GNN-Based Recommender Systems. 20, 1–8.

Guan, H. (2022). An Online Education Course Recommendation Method Based on Knowledge Graphs and Reinforcement Learning. Journal of Circuits, Systems, and Computers, 32(06).

Yao, J., Li, C., Sun, K., Cai, Y., Li, H., Ouyang, W., & Li, H. (2023). NDC-Scene: Boost Monocular 3D Semantic Scene Completion in Normalized Device Coordinates Space. ArXiv (Cornell University).

Barandiaran, X. E., Calleja-López, A., Arnau Monterde, & Romero, C. (2024). A Technopolitical Network for Participatory Democracy: The Future of a Collective Platform. Springer Briefs in Political Science, 119–133.

Yao, J., Wu, T., & Zhang, X. (2023). Improving Depth Gradient Continuity in Transformers: A Comparative Study on Monocular Depth Estimation with CNN. ArXiv (Cornell University).

Yu, Z., Jin, Z., Xue, Q., & Liu, Y. (2023). Comprehensive Teacher Teaching Posture Evaluation Model: Based on OpenPose Posture Recognition in the Context of Digitalization and Intelligence Integration. 479–483.

N. Law, G. Wong, E. Woo, & G. Jiang (Eds.) (2025). *Conference Proceedings of CITERS 2025.*
Hong Kong: The University of Hong Kong.

AI-Agent School: Construction and Evaluation of an Educational Simulation Environment Based on LLM-Driven Agents

Haoming Wang[1], Chengliang Wang[1*] & Chunjia Bao[2]
[1]Faculty of Education, East China Normal University, China
[2]Institute of Systems Science, National University of Singapore, Singapore
[*] wcledutech@163.com

Abstract: This study constructs the AI-Agent School (AAS) simulation environment, aiming to leverage LLM-driven agents to optimize teaching strategies and enhance both teaching and learning outcomes. Additionally, the Zero-Exp strategy is introduced for knowledge base accumulation, significantly improving the capabilities of AI-agents. Experimental results show that students in AAS achieved an overall score of 78.9, outperforming the human teacher control group. This research provides insights and practical references for exploring innovative applications of AI in education.
Keywords: Educational Agents, Teaching Simulation, LLM, Multi-Agent System, Intelligent Education

1. INTRODUCTION

Large Language Model (LLM) agents have demonstrated outstanding performance across various tasks, including code instruction (Jing et al., 2024), information retrieval (Zhu et al., 2023), and answering complex questions (Lin et al., 2024). Among these fields, education is one of the most eager to be reconstructed and reshaped by LLM agents (Wang et al., 2024), as it has a strong demand for personalized adaptive learning and teaching models (Chen et al., 2024). The technical characteristics of LLM agents precisely address this urgent need. However, LLM agents are typically employed to solve specific tasks, such as addressing a particular programming problem in Jing et al. 's (2023) study. There has been no research so far on simulating the entire teaching process, and it remains uncertain whether simulating specific educational scenarios can enhance LLM agents' performance in particular tasks. In light of this, our study will construct AI-Agent School (AAS) based on the logic established by AI Town (Park et al., 2023) and Agent Hospital (Li et al., 2024) to simulate and deduce various teaching and learning processes in schools, thereby forming a corresponding repository of educational experience and knowledge.

2. AI-AGENT SCHOOL SIMULATION ENVIRONMENT DESIGN AND STRATEGIES

2.1. Environment Construction and Role Configuration

AAS is built using Tiled and Cocos (Wang et al., 2024), as shown in Fig. 1, and consists of 25 functional areas that simulate various educational scenarios. The system includes 20 teacher agents, 4 educational researcher agents, and a diverse range of student agents, each with unique backgrounds and characteristics (Fig. 2). AAS simulates six major learning activities: classroom learning, laboratory operations, personalized learning, peer learning, independent learning, and extracurricular activities (as shown in Fig. 3). Teacher agents are responsible for three core tasks: selecting teaching models, applying teaching methods, and formulating personalized guidance strategies.

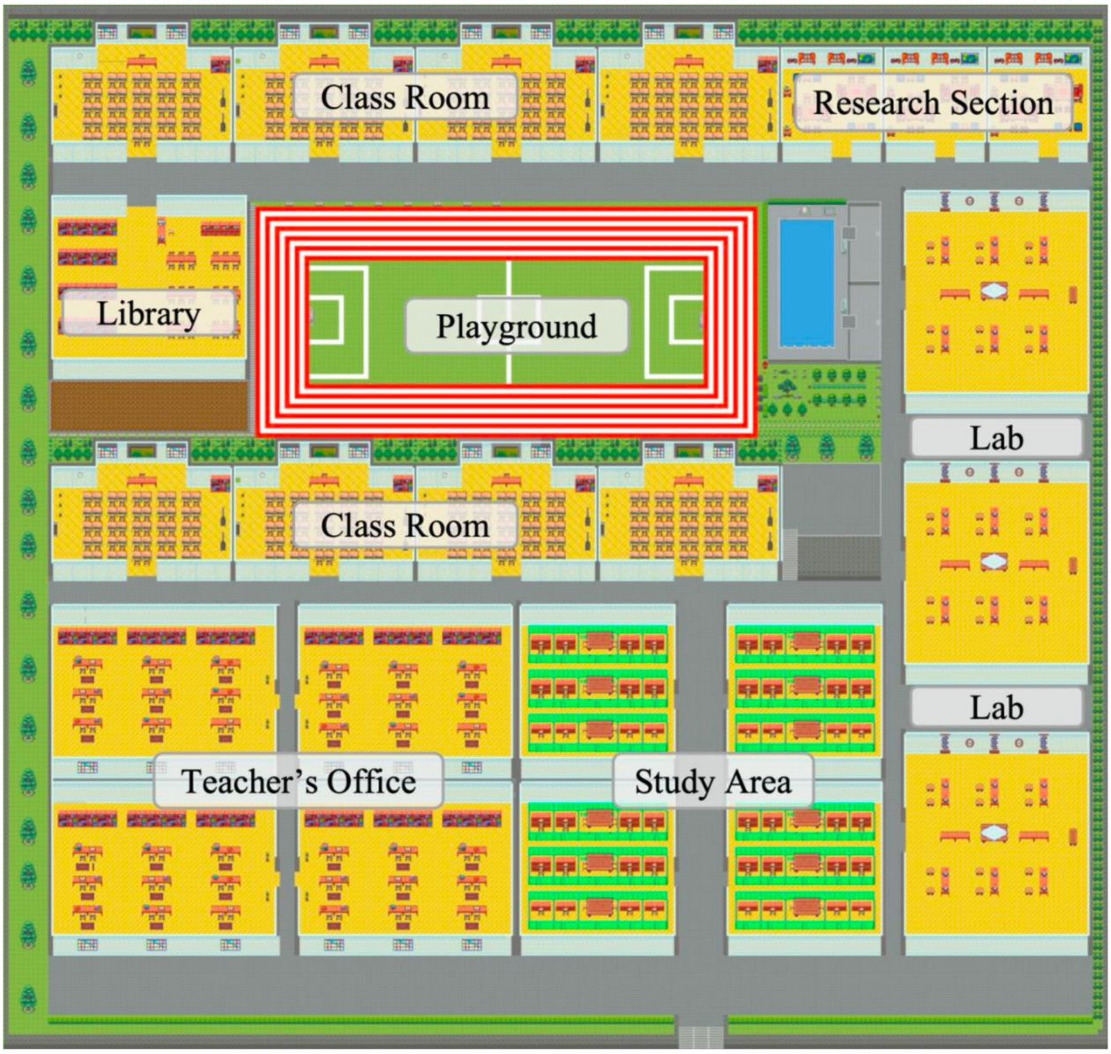


Fig. 1. Structural diagram of AI-Agent school

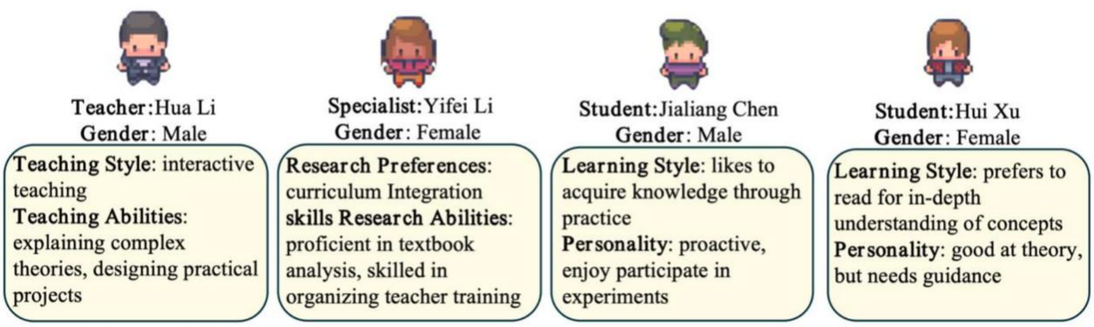


Fig. 2. Partial agent information diagram.

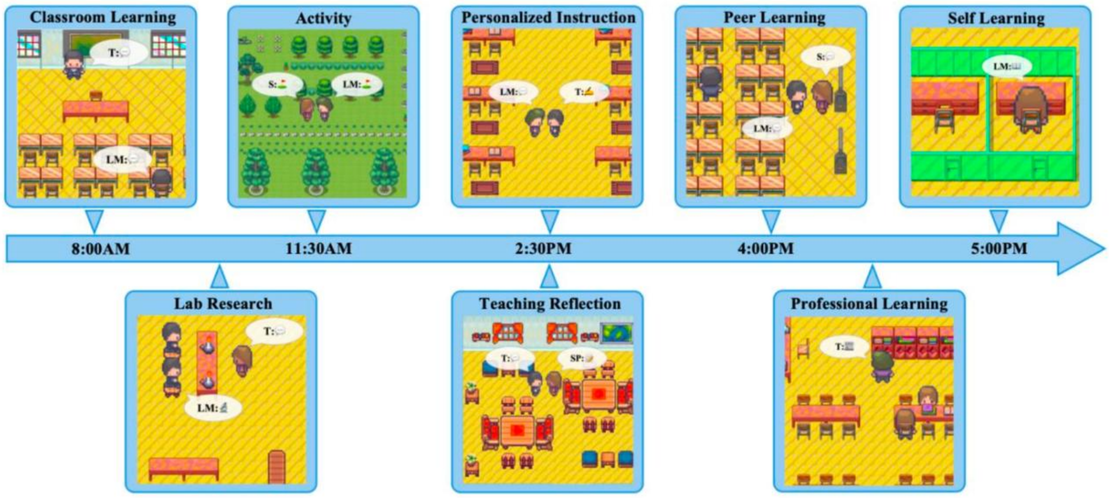
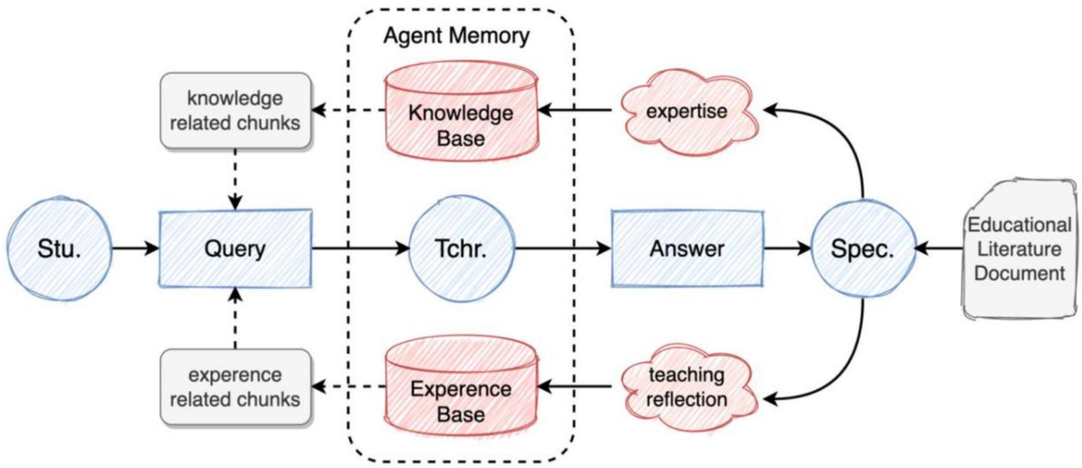


Fig. 3. Activities of Student Agents and Teacher Agents.

2.2. Zero-Exp Strategy and Knowledge Base Construction

This study introduces the Zero-Exp strategy (Fig. 4), leveraging a dynamically updated teaching experience knowledge base and an educational literature knowledge base to enhance the teaching capabilities of teacher agents (Yurtsever et a., 2020). The teaching experience knowledge base is derived from teacher agents' instructional practices within AAS, while the educational literature knowledge base contains approximately 750,000 tokens of academic literature. Additionally, a retrieval-augmented generation mechanism is employed to optimize teacher agents' decision-making in teaching model selection, teaching method application, and personalized guidance strategy formulation.

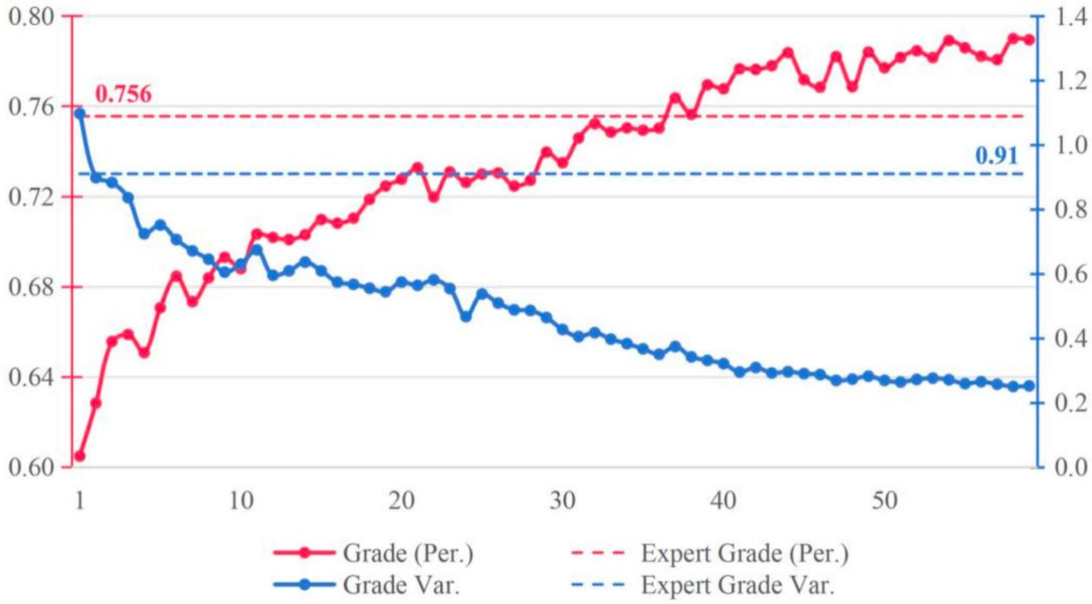
Fig. 4. Zero-Exp Strategy



3. EXPERIMENT SETUP AND RESULTS ANALYSIS

The experiment was conducted with a class size of 30 students over 15 weeks, involving 60 epochs of iterative training. As shown in Fig. 5, the Zero-Exp strategy significantly improved the teaching capabilities of teacher agents. After 60 epochs, the overall class performance of teacher agents reached 78.9%, with a score variance reduced to 0.25, significantly outperforming the human expert control group, which achieved a score of 75.5% with a variance of 0.91. This indicates that AI-agents can rapidly analyze multidimensional class data, optimize teaching strategies, and overcome the cognitive biases and energy limitations of human teachers.

Fig. 5. Performance of Teacher Agents in Teaching Tasks within the AI-Agent School.



The ablation experiment in Fig. 6 confirms the synergistic effect of the teaching experience knowledge base and the educational literature knowledge base, showing that their combined use outperforms their individual usage by 3.27% and 7.08%, respectively. Further analysis reveals that as training iterations increase, teacher agents make more effective use of the knowledge base, leading to a gradual improvement in student agents' knowledge levels (Fig. 7). The test scores at weeks 5, 10, and 15 were 26.3%, 52.8%, and 78.9%, respectively.

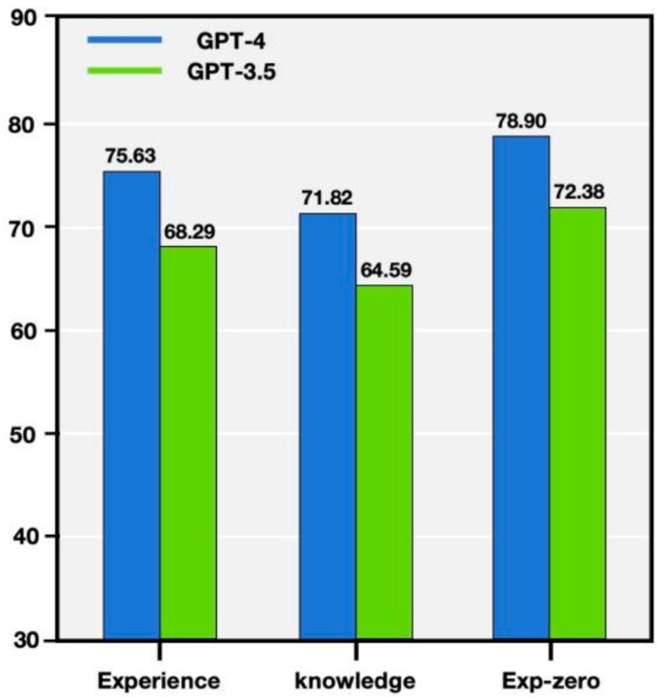


Fig. 6. Ablation Experiment Results of the Zero-Exp Strategy

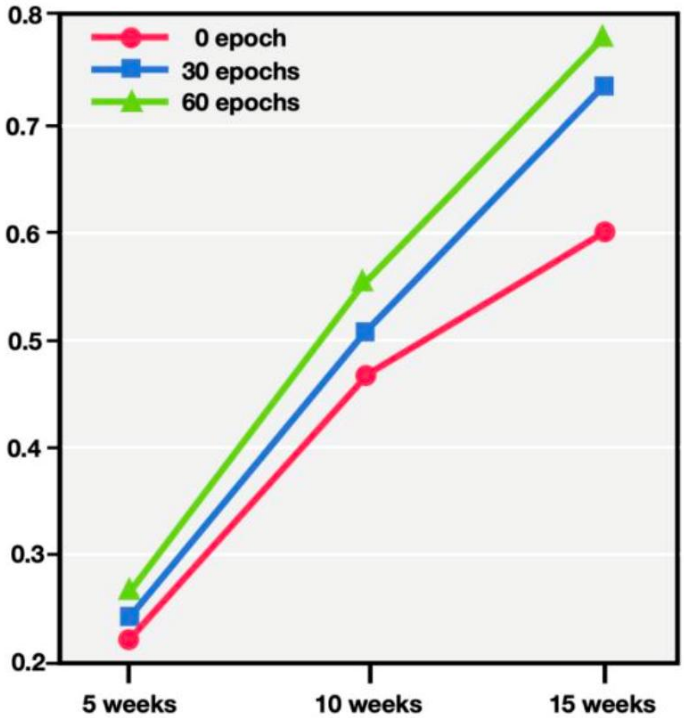


Fig. 7. Knowledge Level of Student Agents in the AI-Agent School.

4. CONCLUSION AND FUTURE PROSPECTS

This study makes significant contributions to both the theoretical and practical aspects of educational technology and intelligent education. From a theoretical perspective, the proposed Zero-Exp strategy and knowledge base integration model not only enrich the design concepts of educational agents but also provide new insights into optimizing teaching strategies in complex educational environments. From a practical standpoint, the study's contributions are equally noteworthy. As a novel paradigm for exploring the application of LLM in education, AAS serves as a valuable reference for advancing intelligent education and expanding new applications of artificial intelligence.

Future research will expand the evaluation of teacher agents to more realistic and diverse educational environments to enhance the external validity of results. We need to focus on the diversity of student agents, including different learning styles, personality traits, and academic levels, to verify the teacher agents' ability to adapt to various learning needs. Meanwhile, the scoring system for evaluating student performance should be improved, establishing multi-dimensional indicators covering knowledge mastery, problem-solving abilities, and more. We recognize the limitations of the current simulation environment, and future research should enhance ecological validity by introducing more complex classroom dynamics and realistic student response patterns. Additionally, we can explore interaction patterns and role distribution between teacher agents and human teachers to optimize teaching effectiveness and promote teacher professional development.

REFERENCES

Jing, Y. H., Wang, H. M., Chen, X. J., & Wang, C. L. (2024). What factors will affect the effectiveness of using ChatGPT to solve programming problems? A quasi-experimental study. Humanities & Social Sciences Communications. 11, 319. https://doi.org/10.1057/s41599-024-02751-w

Zhu, Y., Yuan, H., Wang, S., Liu, J., Liu, W., Deng, C., ... & Wen, J. R. (2023). Large language models for information retrieval: A survey. arXiv preprint arXiv:2308.07107.

Lin, B. Y., Fu, Y., Yang, K., Brahman, F., Huang, S., Bhagavatula, C., ... & Ren, X. (2024). Swiftsage: A generative agent with fast and slow thinking for complex interactive tasks. Advances in Neural Information Processing Systems, 36.

Wang, C. L., Wang, H. M., Li Y. Y., Dai, J., Gu, X. Q. & Yu, T. (2024). Factors Influencing University Students' Behavioral Intention to Use Generative Artificial Intelligence: Integrating the Theory of Planned Behavior and AI Literacy. International Journal of Human-Computer Interaction. 1-23. https://doi.org/10.1080/10447318.2024.2383033

Jing, Y. H., Zhao, L. Y., Zhu, K. K., Wang, H. M., Wang, C. L. & Xia, Q. (2023). Research Landscape of Adaptive Learning in Education: A Bibliometric Study on Research Publications from 2000 to 2022. Sustainability. https://doi.org/10.3390/su15043115

Chen, X. J., Hu, Z. B., & Wang, C. L. (2024). Empowering Education Development through AIGC: A Systematic Literature Review. Education and Information Technologies. 1- 53. https://doi.org/10.1007/s10639-024-12549-7

Li, J., Wang, S., Zhang, M., Li, W., Lai, Y., Kang, X., ... & Liu, Y. (2024). Agent hospital: A simulacrum of hospital with evolvable medical agents. arXiv. https://arxiv.org/abs/2405.02957

Park, J. S., O'Brien, J., Cai, C. J., Morris, M. R., Liang, P., & Bernstein, M. S. (2023, October). Generative agents: Interactive simulacra of human behavior. In Proceedings of the 36th annual acm symposium on user interface software and technology (pp. 1-22). https://doi.org/10.1145/3586183.3606763

Wang, L., Ma, C., Feng, X., Zhang, Z., Yang, H., Zhang, J., ... & Wen, J. (2024). A survey on large language model based autonomous agents. Frontiers of Computer Science, 18(6), 186345.

Yurtsever, E., Lambert, J., Carballo, A., & Takeda, K. (2020). s. IEEE access, 8, 58443-58469.

N. Law, G. Wong, E. Woo, & G. Jiang (Eds.) (2025). *Conference Proceedings of CITERS 2025*.
Hong Kong: The University of Hong Kong.

Empowering High School Students' Self-Regulated Learning in Mathematics Through Structured Reflection Lists and Generative AI

Yanming Wang[1*]
[1]The Second High School Attached to Beijing Normal University, China
[*] yinyue_812@126.com

Abstract: This study explores the integration of structured reflection lists and generative AI tools to enhance self-regulated learning (SRL) in high school mathematics. By combining traditional reflection practices with AI-driven personalized feedback, the research aims to foster metacognitive awareness, problem-solving skills, and student autonomy. Through a qualitative approach involving post-exam reflection, AI interactions, and teacher feedback, the study seeks to evaluate the impact of these strategies on students' mathematical performance and learning behaviors. The initial findings suggest that the combination of structured reflection and AI can promote active self-regulation and improve problem-solving abilities, which offers a novel model for student-centered learning in mathematics education.
Keywords: Self-Regulated Learning (SRL), Generative AI, Structured Reflection, Mathematics Education

1. OBJECTIVES

This study explores how high school mathematics teachers can use structured reflection lists and generative AI tools to enhance self-regulated learning (SRL) in students. By combining traditional reflection practices with AI support, the goal is to improve students' metacognitive awareness, problem-solving skills, and autonomy in learning mathematics.

2. CONCEPTUAL FRAMEWORK

Grounded in Intelligent Learning Design (ILD), this research emphasizes adaptive learning environments tailored to students' needs. ILD integrates personalized learning with technologies like generative AI, which provides real-time feedback and supports self-regulated learning (Callan & Cleary, 2019). Through structured reflection lists, students evaluate their performance, set goals, and refine strategies (Jing et al., 2023). This process, guided by both human and AI feedback, fosters a growth mindset, enhancing students' metacognitive skills and self-regulation (Wanichsan et al., 2021).

3. METHODS

This qualitative study involves the following steps:

(1) Post-Exam Reflection Lists: Students reflect on their exam performance, identifying difficulties and analyzing strategies used (see the attachment).
(2) Generative AI Interaction: After completing the reflection lists, students use AI tools for personalized feedback, with hints, solutions, and targeted exercises.
(3) Teacher Interaction: Students bring unresolved issues to the teacher for further guidance on problem-solving.
(4) Teacher Observation: Teachers assess student engagement, reflection quality, and progress through informal interactions.
(5) Final Exam Performance: Long-term effects will be measured by final exam performance, assessed later.

4. RESULTS

We hypothesize that students in the experimental group will show:

(1) Improved problem-solving skills through personalized AI feedback.
(2) Enhanced self-regulation by adjusting strategies based on AI insights.
(3) Higher academic performance in subsequent exams due to the combined reflection and AI support.

5. SIGNIFICANCE

This study highlights the potential of generative AI to enhance traditional reflection practices, offering a personalized approach to SRL. By integrating AI, the research provides a model for high school teachers to foster more autonomous learners. The findings may inform teaching practices by demonstrating how AI can support reflection and self-regulation, improving academic performance and fostering a culture of autonomy in the classroom.

REFERENCES

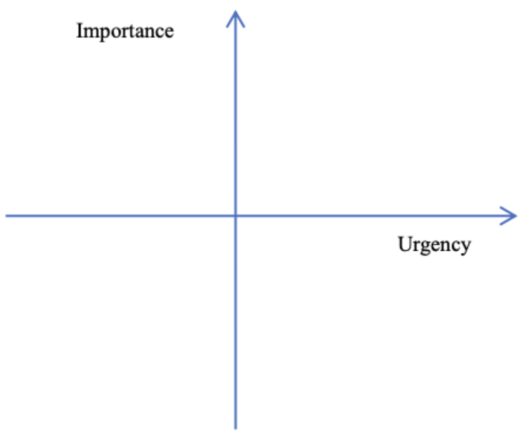
Jing, Y., Zhao, L., Zhu, K., Wang, H., Wang, C., & Xia, Q. (2023). Research landscape of adaptive learning in education: A bibliometric study on research publications from 2000 to 2022. *Sustainability*. *15* (4): 3115.

Wanichsan, D., Panjaburee, P., & Chookaew, S. (2021). Enhancing knowledge integration from multiple experts to guiding personalized learning paths for testing and diagnostic systems. *Computers and Education: Artificial Intelligence*, *2*, 100013.

Callan, G. L., & Cleary, T. J. (2019). Examining cyclical phase relations and predictive influences of self-regulated learning processes on mathematics task performance. *Metacognition and Learning*, *14*, 43-63.

<table>
<tr><td colspan="4">Exam Analysis Appointment Form</td></tr>
<tr><td colspan="4">Basic information</td></tr>
<tr><td>Name</td><td></td><td>Exam name</td><td></td></tr>
<tr><td colspan="4">Self-Description of Exam Performance</td></tr>
<tr><td rowspan="3">Self-Perception</td><td>Overall Impression</td><td colspan="2"></td></tr>
<tr><td>1-2 Points Satisfied</td><td colspan="2"></td></tr>
<tr><td>1-2 Points Dissatisfied</td><td colspan="2"></td></tr>
<tr><td rowspan="6">Detailed Statement
Please use questions from this exam as examples</td><td>Mastery of Basic Concepts
(Please explain in detail if there are any issues)</td><td colspan="2"></td></tr>
<tr><td>Mastery of Basic Methods
(Please explain in detail if there are any issues)</td><td colspan="2"></td></tr>
<tr><td>Writing, Process, and Steps
(Please explain in detail if there are any issues)</td><td colspan="2"></td></tr>
<tr><td>Problem-Solving Strategies
(Please explain in detail if there are any issues)</td><td colspan="2"></td></tr>
<tr><td>Mental Condition</td><td colspan="2"></td></tr>
<tr><td>other</td><td colspan="2"></td></tr>
</table>

<table>
<tr><td colspan="3">Previous Learning Situation</td></tr>
<tr><td>Consistent Study Habits and Self-Evaluation</td><td colspan="2"></td></tr>
<tr><td>Current Study Strategies and Self-Evaluation</td><td colspan="2"></td></tr>
<tr><td colspan="3">Purpose and Expectations of the Meeting</td></tr>
<tr><td>Current Problem Description
(Please identify the problem in the corresponding quadrant)

Note:
1. "Urgency" is represented from left to right, from weak to strong.
2. "Importance" is represented from bottom to top, from weak to strong.
For example, the first quadrant represents problems that are both important and urgent.</td><td colspan="2">
</td></tr>
<tr><td colspan="3">Problem-Solving Strategies</td></tr>
<tr><td>Problems That Can Be Self-Solved</td><td>Problems Solved Through Peer/AI Suggestions</td><td>Problems for Which Teacher Advice Is Needed</td></tr>
<tr><td></td><td></td><td></td></tr>
</table>

N. Law, G. Wong, E. Woo, & G. Jiang (Eds.) (2025). *Conference Proceedings of CITERS 2025.*
Hong Kong: The University of Hong Kong.

A Study on the Characteristics of Undergraduates' Learning Grit in Online Collaborative Learning Peer Feedback

Sijie Zhang[1*], Minjie Zhu[1], Xiaoran Chen[1], Xinyan Ye[1], Heting Li[1] and Xilin Wu[1]
[1]School of Information Technology in Education, South China Normal University, China
[*] zhangsijie1003@163.com

Abstract: Current research has indicated that resilience, grit and growth mindset are psychological characteristics beneficial to learning. This paper addresses a significant gap in the literature, that being that little research has connected the concepts of grit and peer feedback to online learners, and to elements specific to the collaborative learning environment. The study was conducted over a seven-week duration, utilizing an online course on Educational Communication taught to 34 second-year undergraduates at the University of H. Students' grit were assessed using a pre- and post-test grit dispositions questionnaire. Additionally, lagged sequence analysis and R studio were employed to analyze the peer feedback provided by students. The findings of this study reveal that peer feedback indeed promotes grit among undergraduate students. Furthermore, distinct patterns of peer feedback behavior were identified, each having a unique impact on students' grit. These findings provide valuable insights for instructors seeking to effectively leverage peer feedback in fostering grit among students in subsequent online teaching environments.
Keywords: peer feedback, grit, perseverance of effort, consistency of interest, collaborative learning

1. INTRODUCTION

Higher education is increasingly promoting student learning by conducting online collaborative learning activities. Learning PERSEVERANCE appears important for students in digital environments, which can have a direct impact on academic success and attrition rates (Barber et al., 2019) . Learning perseverance is defined as the actions of students to persevere in the face of learning difficulties or problems, and it is an important factor affecting the learning effect of students (Farrington et al., 2012). Research shows that a way to increase perseverance in learning with manipulatives may be to foster collaborative learning (Carbonneau et al., 2020). The advantages associated with collaborative learning include enhanced student engagement, increased learning achievement, development of higher-order thinking skills, abilities to solve problems in novel ways (Johnson et al., 1991), and increased student perseverance (Loes et al., 2017). The opportunity to consider multiple perspectives through collaborative learning may enhance learners' positive attitude toward learning and increase the likelihood that they persevere on hard tasks longer. The results from Saitta et al. (2011) provide initial support for this statement.

There is growing evidence which suggests that peers' ideas (e.g., feedback, project outcomes) can stimulate their fellow learners' creativity in contexts of interaction (Pi et al., 2019a; Pi et al., 2019b; Ziegler et al., 2000). One common way to promote interaction between learners is through peer assessment (Çevik, 2015). Peer assessment refers to a process in which learners review peers' performances or produced artifacts (Topping et al., 2000). Previous studies have suggested that peer assessment is more beneficial in online settings as compared to traditional settings (Tsai & Liang, 2009; Yang & Tsai, 2010). Through online peer assessment, learners can engage in reviewing peers' works and process them on a deeper level by providing feedback, while also being able to better assess their own performance through reading the feedback given to them by their peers. Through activities such as this, learners achieve a greater understanding of the topic, engage in more thought, and therefore come up with more creative ideas (Honeychurch et al., 2013; Pi et al., 2019a; Pi et al., 2019b; Ziegler et al., 2000), however, no research has focused on the refinement of online peer feedback scripts to promote student persistence.

2. LITERATURE REVIEW

2.1. Grit in online collaborative learning

The growth of online collaborative learning and the ensuing environmental shift in education to include digital learning platforms has caused educators to look closely at the way students best experience and online collaborative learning (Weegar & Pacis, 2012). Smart and Cappel (2016) found online learning brought with it time pressures for completing modules, and provided inadequate opportunity for human interaction, a factor deemed necessary for establishing peer support and developing in-depth group discussions on subject matter. Other pressures were identified including learner motivation, time management, comfort level with online technologies, as well as technical problems, a perceived lack of sense of community, time constraints, and difficulties understanding online course objectives (Barber et al.,2019; Goegan et al., 2023).

In order to address these issues, grit, as a critical notion, has drawn the attention of educational investigators (Gray & Mannahan, 2017). According to Duckworth (2016), grit, as a non-cognitive construct, is more important than talent in academic achievement. Cross (2014) described grit as tolerating difficulties though preserving the wish for long-term purposes. Von Culin et al. (2014) argued that gritty learners have willpower, perseverance, capability to determine well-defined objectives, endurance, and flexibility in facing difficulties in their academic life. This study suggests that grit in online collaborative learning refers to the perseverance and passion for long-term goals despite obstacles encountered in virtual group work and communication settings.

Online collaborative learning in higher education has expanded considerably following the restrictions imposed during the pandemic (Hernández-Sellés et al., 2023). Few studies have been done on the role of grit in learners' performance in online educational contexts. For instance, Aparicio et al. (2017), in their study, found out that learners' grit in online learning is significantly correlated with their satisfaction and performance. Malureanu et al. (2021) explored the relationship among the individuals' beliefs regarding self-confidence, grit, ease of use, self-efficacy, and usefulness of eLearning platforms during the COVID-19 pandemic. They found out that grit, self-efficacy, and perceived ease of use of eLearning platforms were considerably directly influenced by the self-confidence variable. Mediation analysis indicated that full mediation occurs only through the ease of use of eLearning platforms variable in the relationship between self-confidence and usefulness. Yang (2021) explored the relationship between Chinese EFL students' grit, wellbeing, and classroom enjoyment in online classrooms during the COVID-19 pandemic. They found out that there is a positive relationship between learners' grit and enjoyment, and high degrees of enjoyment were interrelated to high degrees of grit. They found out that grit significantly predicted students' wellbeing and was also a predictor of classroom enjoyment. Zhao's(2022) review aimed at exploring the related investigations on the effects of online and traditional learning contexts on English as a foreign language learners' grit and foreign language anxiety, have verified the relationship between learners' grit and academic performance in online learning contexts. It can be seen that grit is an important factor affecting the effect of college students' online collaborative learning. Few studies have explored how to improve grit in online collaborative learning.

2.2. Online peer feedback script

In recent years, a variety of online environments have been designed to support peer feedback processes and its outcomes (Latifi et al., 2019; Latifi et al., 2020). The online peer feedback environments allow students to submit their works, provide feedback to their peers' works reciprocally and anonymously and continuously revise their works based on feedback received from their peers, without restriction of time and space (Tsai, 2009). Also, such environments increase timeliness of feedback for learners to reflect on one's own and their peers' work (Chen & Tsai, 2009). Next to these benefits for students, online peer feedback environments also have notable benefits for instructors. For example, these environments enable instructors to systematically manage peer feedback processes and monitor progress of and interactivity between students (Chen & Tsai, 2009). Online environments can increase validity and reliability of peer feedback by using anonymous online peer feedback and clarification of criteria embedded in the systems for example in terms of rubrics (Wen & Tsai, 2008). Implementing peer feedback process in online environments could also decline instructor's workload (Davies, 2000) and peer feedback time (McGourty, 2000). Using online peer feedback, instructors can automatically collect and record data about students activities such as time spent on task, off-tasks, degree of participation and communication among learners, and use them for further learning analytics (Tsai, 2009). With advancement and possibilities of online learning environments in the recent years, various approaches, such as worked example and scripting strategies can be offered to support students with providing high- quality peer argumentative feedback in these environments.

Scripting is a typical instructional approach that has been frequently used to scaffold various aspects of the learning processes and outcomes in a more active form than worked examples (Gan & Hattie, 2014). Scripts are seen as specific type of scaffolds in form of detailed and explicit guidelines or instructions that help students engage in a structured and desired learning processes to achieve expected learning outcomes (Kollar et al., 2006). Recent studies revealed that providing structure is essential to support learners in generating high-quality peer feedback (Peters et al., 2017). Students who receive and/or provide high quality peer feedback, often produces a high level of learning outcomes (Noroozi et al., 2016). Core elements of scripts are prompts that cue students on how to identify weaknesses and strengths in a learning product and to generate specific suggestions for improvement (Peters et al., 2017). It can be seen that online peer feedback scripts proposed by existing studies are mostly used to improve the quality of written argumentative essays, but do not focus on the diversified peer interaction behaviors in online collaborative learning.

2.3. The effect of peer feedback on grit

At present, the research on peer and perseverance shows that peer relationships have an impact on learning perseverance. For example, a research examined the moderating role of developmental stage (adolescence vs. emerging adulthood) in the association between peer attachment and learning perseverance. The results revealed that peer attachment contributes more to perseverance of effort in emerging adulthood (Lan, 2017). Jin et al.'s research shows that perseverance plays a partial intermediary role between peer attachment and academic procrastination; More precisely, peer attachment is positively correlated with perseverance, while perseverance is negatively correlated with academic procrastination (Jin et al., 2017). So will peer feedback have an impact on learning perseverance? At present, it is still an open question. Therefore, in this study, it is analyzed whether learning perseverance will be affected under the peer feedback strategy, so as to expand the understanding of the interaction between learning perseverance and peer feedback. Moreover, few studies have studied the behavior patterns or sequences of students with different perseverance levels. Therefore, in this study, in addition to comparing academic performance, we also try to find and compare the behavioral characteristics of students with different perseverance levels.

3. RESEARCH QUESTIONS

Above all, the current research on peer and grit shows that peer relationships, such as peer attachment and peer association, have an impact on learning grit. However, how the peer effect interacts with non-cognitive traits (that is, learning grit) is still an open question. Therefore, in this study, it is analyzed whether learning grit will be affected under the peer feedback strategy, so as to expand the understanding of the interaction between learning grit and peer feedback. Moreover, few studies have studied the behavior patterns or sequences of students with different grit levels. Therefore, in this study, in addition to comparing academic performance, we also attempt to find and compare the behavioral characteristics of students with different grit levels and answer the following three questions:

(1) Do scripted online peer feedback activities improve grit among undergraduates?
(2) Is there a relationship between changes in undergraduate grit scores and undergraduate engagement?
(3) If so, which transitional patterns of peer feedback behavior are associated with increased grit?

4. METHODOLOGY

4.1. Research design and data collection

The study was conducted in a course titled Educational Communication in the fall semester of 2022 in the Educational Technology program at a university in southern China. The participants were second year university students in the course, 34 in total of which 9 were male and 25 were female, with a mean age of 20 years (SD = 0.45). All signed a consent form agreeing to the use of their data for research purposes. The 16-week program was delivered in a blended format that included face-to-face and online courses as well as out-of-class online learning activities (e.g.,online discussions and viewing lecture videos). The study was conducted during weeks 6 through 12 (Figure 1).

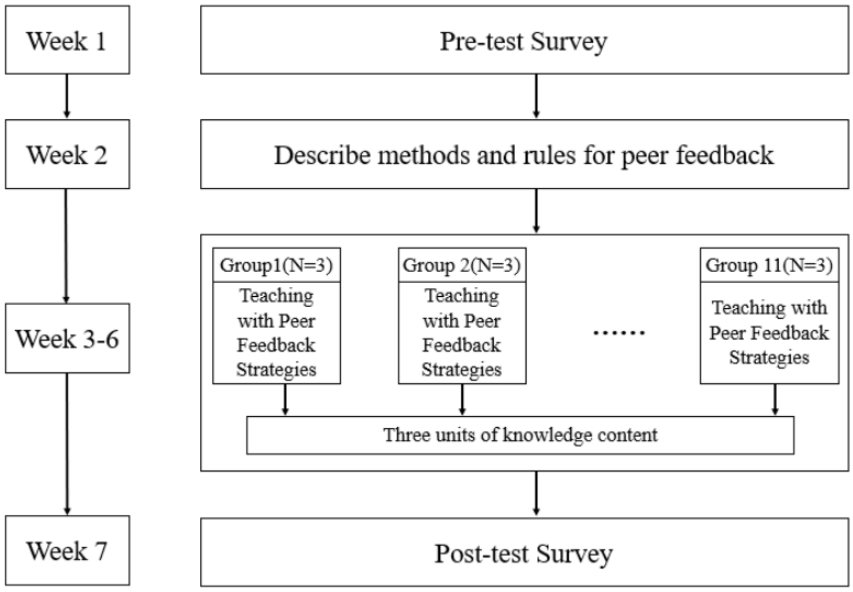


Fig. 1. Experimental Procedure Diagram

Prior to the start of the study, students were randomly assigned to 11 groups of three to four students each to optimize the benefits of collaborative learning. Sessions during the first and second weeks of the study were conducted in a face-to-face format. During the first week, teaching assistants explained the collaborative task, which was to create a micro-course, to the participants. To help students engage in collaborative learning through peer feedback behaviors, the TA explained the rules of effective peer feedback and provided examples and a presentation template (e.g., I agree/disagree with the statement you mentioned in section xxx for xxx). At the same time, a questionnaire on the Pre-test of Grit Scale was distributed. After consulting with experts in the field of grit, five items were selected from Duckworth & Quinn's (2009) Short Grit Scale, which is based on Duckworth et al. (2007) Grit Scale. Each item had five options ranging from 1 = very non-compliant to 5 = very compliant. The questionnaire was piloted with 21 sophomore students at another university. Based on the results of the pilot test, we modified some of the questions that were prone to misinterpretation and ambiguity. Table 1 shows the alpha coefficients of the final questionnaire and the model fit indices of the confirmatory factor analysis in the sample of this study. These statistics indicate that the internal consistency and structural validity of the questionnaire are acceptable.

Table. 1 The internal consistency and structural validity of the questionnaire

	Coefficient α	χ^2/df ***(p)***	**CFI**	**TLI**	**RMSEA**	**SRMR**
Pre-test	0.95	0.981(0.08)	0.934	0.867	0.258	0.046
Post-test	0.95	0.677(0.08)	0.976	0.952	0.112	0.043

In weeks 2 to 5, the classes consisted of lectures and group discussions (about 20 minutes per session). During the discussions, students focused on different parts of the assignment, such as planning a micro-course and designing interactive activities. Students were asked to contribute at least once to each discussion. The rest of the work of creating the micro-course was done asynchronously outside of class. In week 6, groups presented their work in class and received advice from teaching assistants, based on which students discussed how to improve their work. In week 7, students complete the post-test Grit Scale.

4.2. Data categorization and analysis

A coding scheme (see Table 2) for peer feedback types was developed based on the peer feedback rating scale proposed by Burgess et al. (2021). This scheme was tailored to align with the specific course activities and experimental design. We refined the original items and transformed them into codes that represented distinct feedback behaviors. The final coding scheme included five types of peer feedback behaviors and nine behavior codes.

Data comprised group conversation and the Grit Scale questionnaire's pre-test and post-test results. After removing data that could not be compared to the pre-test and post-test, a total of 35 pre-test scale data and 34 post-test scale data were found, yielding a scale validity of 0.97. In the group discussion, every round was seen as an example of a peer feedback behavior. A total of 4879 peer feedback behaviors were observed online, with an average of 444 activities per group. The feedback behaviors were evaluated separately by two experienced coders, yielding a kappa value of 0.85 and a concordance rate of 0.86. This degree of agreement is suitable. When the developers couldn't agree at first, they worked things out and came to an agreement on the finished product.

Table. 2 Peer Feedback Behavioral Coding Scheme

Type of behavior	Code	Behavior	Example
Cognitive behavior	A	Listen to and respect others' opinions	"That's right, put it together, just the last part of the word cloud put it together into a title."
	B	Encourage, endorse, and evaluate others' opinions	"You're right, it's important to be clear about what we're about first. But the first step is to determine the topic."
Task solving behavior	C	Work closely with team members	"We've just discussed the elements of a micro-course title, so is there anything else you'd like to bring up?"
	D	State confusion and questions about the material and opinions related to the assignment	"That's you up there eh, what software are you using now?"
Task breakthrough behavior	E	Articulate key ideas and opinions facilitating problem-solving and breakthroughs	"Not always find some symbols that fit very well. Too many elements piled on top of each other leads to a theme where it's easy to jumble everything together and not catch the point."
	F	Creatively solve issues in collaboration	"We discussed AI assisting us in creating a micro-course, so can we use a robot as a micro-course presenter?"
Corrective behavior	G	Indicate or correct problems in group work	"The term technology feels a little too broad."
	H	Accept criticism and feedback and respond appropriately	"During the debriefing, the students said that our knowledge points were presented in a rather abstract way and that we could introduce them in more detail when we revise them."
Results-testing behavior	I	Reflective feedback after group work presentation (from the perspective of optimizing the work)	"It could be here that their 9th group says they can add click-through interactive content, and I think it would be best if there was click-through interactive content."

In order to determine if students' grit changed as a result of the online peer feedback activity, a paired-sample t-test was used for Q1.

For Q2, the relationship between the posterior scores and undergraduate engagement was discussed by controlling for undergraduate grit in learning pre-test scores using a partial correlation analysis. This was done in order to investigate whether changes in undergraduates' post-test scores were related to undergraduate engagement (total number of times they participated in discussions (DT), total number of times each behavior occurred).

For Q3, a typical analytic procedure for examining associations between behavioral shifts and external variables (e.g., achievement, motivation level, domain knowledge) consists of discretizing the external variables into categories (e.g., high versus low levels) and comparing each behavioral shift using statistical tests. However, this approach has two limitations. First, converting numeric variables to categorical variables may reduce statistical power and lead to biased results (Dawson & Weiss, 2012; Maxwell & Delaney, 1993). Second, there may be many behavioral transitions (e.g., 9*9 = 81 transitions in this study). Testing each transition may increase the Type I error rate.

To address Q3, we devised a novel analytical methodology integrating lag-sequential analysis, clustering techniques, and repeated ANOVA. Precisely, we derived the transition matrix for each group, quantifying the synchronous peer feedback behaviors. Each matrix cell encapsulated the log-odds ratio of a behavior from one row transitioning to a behavior in the corresponding column, serving as an indicator of the strength of this transition. Notably, we eschewed the classic LSA's z-score or adjusted residuals due to their inadequacy in measuring transition strength, as underscored by Bakeman & Quera (2011). Instead, log-odds ratios were preferred for their ability to discern group-wise differences in transition strength. In this context, the metric signified the odds ratio of a behavior in the row transitioning to another specified in the column, relative to transitioning to any behavior not in that column. Subsequently, we leveraged the transition matrices to construct a distance matrix by summing the disparities between each group pair's matrices. The cell entries in this matrix reflected the aggregate differences between the respective groups. We then employed Ward's clustering algorithm on the distance matrix, leveraging its efficacy in grouping students based on their behavioral sequences, as evidenced by Ouyang et al. (2023). Groups within the same cluster exhibited greater homogeneity in their peer feedback transition patterns compared to inter-cluster groups. Concluding our analysis, we employed a two-way repeated ANOVA to investigate the moderating effect of cluster membership on grit evolution. Here, grit scores served as the dependent variable, time as the intra-subject factor, and cluster membership as the inter-subject factor. This approach enabled us to discern whether belonging to a particular cluster influenced the evolution of grit over time.

5. RESULTS

5.1. Q1: Do scripted online peer feedback activities improve grit among undergraduates?

The results of the paired-samples t-test indicated a statistically significant increase in students' grit (see Table 3). The change in the mean score for each item was 0.34 with a large effect size (Cohen's d = 0.81).

Table 3. Pre-test Post-test Paired Samples Statistical Table

Time	Median	Mean	*SD*	*t*	*p*	*Cohen's d*
Pre	4	4.08	0.76	2.41	0.02	0.81
Post	4.3	4.28	0.57			

5.2. Q2: Is there a relationship between changes in undergraduate grit scores and undergraduate engagement?

In the case of pre-test grit scores, a partial correlation was used to determine the relationship between undergraduates' grit post-test scores and the undergraduates' level of participation in the discussion (number of times the undergraduate participated in the discussion and number of times the undergraduate switched each behaviour). In the case of the pre-test grit score (4.08 ± 0.13), the partial correlation between the undergraduates' grit post-test score (4.28 ± 0.57) and the total number of times the undergraduate participated in the discussion (78.74 ± 6.76) was not statistically significant, $r(31) = 0.18$, $N = 34$, $p = 0.32$. Meanwhile, there was a statistically significant relationship between the undergraduates' grit post-test score and the number of times the undergraduate switched each behaviour. There was no statistically significant correlation between the number of times a undergraduate switched behaviours (see Figure 3), which suggests that the number of times a student switched behaviours per behaviour did not have a significant effect on the undergraduates' grit post-test score. It can be assumed that there is no correlation between the level of student participation in the discussion and grit, and in order to further explore whether there may be a correlation between sequence switching patterns and grit, we developed an analysis of Q3.

Table 4. The Partial Correlation Between Post-test Grit Scores and the Total Number of Times Undergraduates Participated in the Discussion(DT) Table

Observational item		Post-test	DT
Post-test	Correlation	1.00	0.09
	Significance (2-tailed)	.	0.62
	df	0	31

5.3. Q3: If so, which transitional patterns of peer feedback behavior are associated with increased grit?

Figure 3 displays the dendrogram of Ward's clustering, with a cophenetic correlation coefficient of 0.82, suggesting that the dendrogram well represented the distances between groups' peer feedback transitions. According to this dendrogram, there were two clusters of groups with distinct feedback patterns. Groups G1, G2, and G5 belonged to one cluster (C1), while the remaining groups constituted another cluster (C2).

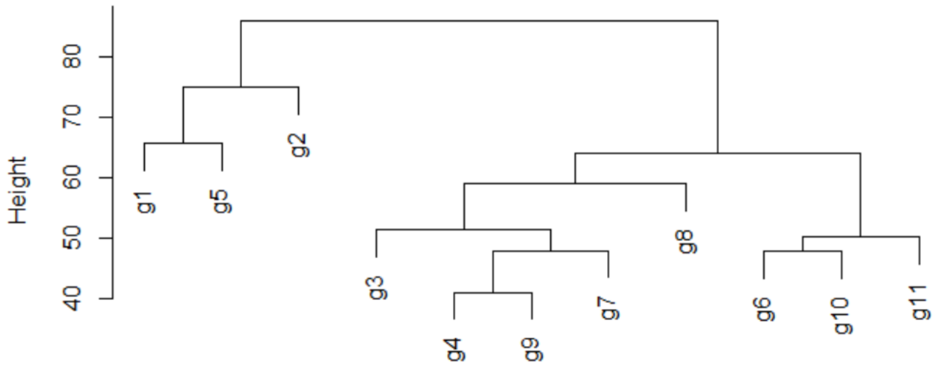


Fig. 2. The dendrogram of ward clustering based on the similarities in the patterns of online synchronous peer feedback behaviors

Repeated ANOVA was conducted to investigate if changes in grit scores differed by clusters (see Table 6). The results indicated no significant main effect for the cluster ($p = 0.40$, partial $\eta 2 = 0.02$); no significant main effect for time ($p = 0.11$, partial $\eta 2 = 0.02$); and a significant interaction between cluster and time at the 0.05 significance level ($p = 0.03$, partial $\eta 2 = 0.04$). The simple main effect test indicated that grit scores were significantly higher in the post-test than the pre-test in cluster C2, with a mean value increase of 0.36 ($p = 0.01$; Cohen's $d = 0.36$; see Table 7). However, such a difference was not found in cluster C1 ($p = 0.45$; Cohen's $d = 0.24$), and the mean value decreased by 0.24.

Table 5. Repeated ANOVA tests for within-group and between-group variables

		SS	*df*	MS	F	*p*	η^2
Between	Cluster	0.48	1	0.48	0.73	0.40	0.02
	Residual	21.02	32	0.66			
Within	Time	0.68	1	0.68	2.74	0.11	0.02
	Cluster*Time	1.21	1	1.21	4.87	0.03	0.04
	Residual	8.00	32	0.25			

Table 6. Two paired-sample t-test results for undergraduate pre-and post-tests of two online collaboration peer feedback behavioral models

Cluster	Pre-test/Post-test	*SD*	*t*	*p*	*Cohen's d*	Mean	Sum
C1	Pre	0.84	-0.57	0.45	0.24	4.44	40
	Post	0.63				4.20	38
C2	Pre	0.72	1.88	0.01	0.36	3.95	99
	Post	0.56				4.31	108

Note: To investigate which transitions of peer feedback behaviors might result in the differences in the change of grit scores, we compared the transition matrices between the two clusters (Tables 6 and 7). We focused on transitions that were relatively strong in at least one cluster and were substantially different between C1 and C2 (Rosenthal, 1996)[1]. In this study, we consider that odds ratio larger than 2.0 has a relatively high probability of occurrence. Comparing C1 and C2 (see Tables 8 and 9), excluding the I-I and G-H behavioral transitions, which occurred at a larger rate in both groups, the five behavioral transitions F-C, E-F, D-H, H-H, and H-I occurred at a larger rate in C1.E-F and F-C indicated that articulating key ideas and opinions crucial for problem-solving and breakthroughs fostered creative collaborative problem-solving, which in turn promoted closer teamwork. D-H, H-H, and H-I indicated that when one group member gave feedback on his/her confusion and questions about certain materials/opinions in the task, and the other member accepted the criticism and feedback and responded appropriately, it could elicit reflective behaviors from the group members in response to the debriefed work.

In C2, three types of transitions, F-A, C-C, and A-G, occurred at relatively high rates. In C2, the three transitions F-A, C-C, and A-G occurred at a relatively high rate. F-A and A-G indicated that creative problem solving in group collaboration provokes students to listen to and respect the opinions of their group members, which in turn lead to corrections of what is not right in the work. C-C indicated that groups likely stayed in a state of close collaboration.

Table7. C1 Table of coding indices

	A	B	C	D	E	F	G	H	I
A	0.50	1.88	0.50	1.85	0.90	0.67	1.61	1.17	0.98
B	0.66	0.72	1.77	0.73	1.79	0.62	0.85	1.13	0.99
C	1.18	0.91	0.67	1.82	1.35	0.80	1.01	0.59	0.52
D	0.39	0.87	0.97	0.43	1.50	1.33	1.21	6.24	0.73
E	1.85	0.81	1.05	1.19	0.58	2.88	1.09	0.21	0.82
F	1.43	1.92	3.42	0.93	0.10	0.51	1.43	0.41	0.78
G	0.42	0.98	0.30	1.62	1.73	1.70	0.57	2.62	1.39
H	0.64	1.14	0.92	1.65	1.10	1.41	0.89	2.38	2.18
I	1.28	0.47	1.51	0.59	1.06	0.51	1.14	0.51	4.18

Note. Cell values are the odds ratio of the row code transitioning into the column code. For, the value of the cell in row F and column C is 3.42, which means that the ratio of the odds of behavior F followed by behavior C to the odds of behavior F followed by behaviors other than C is 3.42. This ratio suggests a strong transition strength from behaviors F to C.

Table 8. C2 Table of coding indices

	A	B	C	D	E	F	G	H	I
A	0.21	0.86	0.95	1.14	0.97	1.59	2.09	1.95	1.36
B	0.60	0.78	1.08	1.04	1.43	1.41	0.85	0.76	0.79
C	1.16	0.44	3.07	1.02	0.97	0.93	0.58	0.38	0.46
D	1.11	0.89	0.79	0.27	1.80	0.91	0.41	2.70	0.70
E	1.26	1.55	0.67	1.49	0.70	0.94	1.51	0.33	0.70
F	2.33	1.65	0.72	1.01	0.55	1.17	1.60	0.39	1.47
G	1.69	0.69	0.53	1.08	0.62	1.20	0.77	3.36	1.42
H	1.42	1.19	0.73	1.40	0.83	0.66	1.31	0.40	1.39
I	0.82	1.95	0.66	1.01	0.85	0.54	0.85	0.69	5.05

6. DISCUSSION

6.1. The Change in Grit Following Peer Feedback Activities in Online Collaboration

The results of Q1 indicate that online collaborative peer feedback activities improve undergraduate students' grit, which is consistent with previous studies (e.g., Sun et al., 2024; Howell et al., 2018). These findings imply that online collaborative peer feedback activities are effective in promoting the development of grit in undergraduates. The results of Q2 provide insights into the mechanisms underlying this effect.

6.2. C2 Online Collaborative Peer Feedback Behavior Patterns Changes in Grit
Three primary behavioral transitions (C-C, F-A, and A-G) in the C2 group's behavioral transition pattern had a greater likelihood of happening differently from C1. After determining that the C2 group had substantially higher grit scores in the two paired samples t-test, we closely examined the three behavioral patterns.

The C-C transition represented a continuum of behaviours that work closely together within a team (e.g., "Let's start by talking about non-verbal symbols. I'll work on that first. – The micro-course has shown that emotions can be expressed and tone of voice can be reinforced'). This shift in feedback behaviour can promote an atmosphere of cooperation and interaction. According to Lee et al. (2019), undergraduates' grit promoted problem solving in a collaborative learning environment. The results of this study suggested that when designing instructional strategies to develop undergraduate grit, educators should consider whether activities promote support and collaboration among undergraduates. The F-A, A-G transitions indicated that undergraduates engage in active listening in order to facilitate the resolution of group collaboration issues, subsequently identifying areas for improvement in the group's work (e.g., 'I think the opening credits are too long at the moment' - 'Do you have any suggestions for changes?' - 'I think the import section could be compressed in length to less than 30 seconds.'). Ekwonye et al. (2023) state that in collaborative learning, undergraduates' grit encourages peers to solve key problems, and issue resolution inadvertently somewhat fosters undergraduates' grit. Because of this, while creating educational activities, teachers may intentionally assist students in critically collaborating on issue solving.

6.3. C1 Online Collaborative Peer Feedback Behavior Patterns Changes in Grit
Though the C1 group's grit scores were significantly lower in the grit pre and post-tests, it can be argued that these five behavioral transition patterns are not supportive of undergraduates' development of grit. The F-C, E-F, D-H, H-H, and H-I transitions were specific to the C1 group, and they did not significantly affect the development of undergraduates' grit.

For example, the E-F, F-C transitions imply that undergraduates express themselves clearly in group work, and creative problem-solving triggers undergraduate behaviours that work closely with team members and may promote undergraduates' cooperative communication skills. However, none of the transitions in this study were associated with learning grit. Therefore, there is no need to support these transitions in peer feedback activities when the goal is to promote learning grit. The D-H, H-H, and H-I transitions imply that undergraduates' feedback of their confusion and questions about certain points in a collaborative group task triggers the behaviour of receiving criticism and feedback from members and responding appropriately, which in turn triggers the behaviour of reflective feedback from members after debriefing on the group's task work. However, in this study, these behaviours may lead to a situation where the goals achieved deviate from the undergraduates' stated goals, which in turn leads to a reduction in learning grit as undergraduates overcome difficulties and solve the task without achieving the mentally expected goals. Therefore, teachers can guide members to encourage each other to face challenges and re-orientate goals flexibly with a positive mindset and improve self-adaptation in the design of teaching and learning activities (Morton & Paul, 2019).

7. CONCLUSIONS, LIMITATIONS AND FUTURE DIRECTIONS

The present study demonstrates that online collaborative peer feedback activities can positively influence undergraduates' grit development by fostering cooperative behaviors and sustained engagement with long-term goals. While the behavioral transition patterns observed in the C2 group (e.g., C-C, F-A, A-G) were associated with grit enhancement through collaborative dynamics, the patterns in the C1 group (e.g., F-C, E-F, H-H) highlighted the limitations of task-focused communication alone. Despite methodological constraints such as the lack of a control group, small sample size, and short study duration, the findings suggest that educators should prioritize designing peer feedback activities that promote teamwork, reflective practice, and incremental challenge to cultivate grit.
To address the needs of students with low grit, educators should implement strategies focused on building psychological resilience and effective learning habits. First, reframing confusion as a catalyst for growth can be achieved by introducing the concept of the "optimal confusion zone," helping students perceive challenges as opportunities for learning. This approach enhances psychological resilience and curiosity. Second, structured self-regulation practices, such as setting phased learning goals, time management training, and task decomposition, can reduce overwhelm and improve self-efficacy. Third, reflective journaling can transform frustration into actionable insights by encouraging students to document and analyze setbacks. Finally, fostering social support networks through peer mentorship and teacher-student collaboration creates a safety net for high-confusion scenarios, emphasizing communication and collective problem-solving.

For educators, dynamic task design is critical. Gradually calibrating task difficulty based on students' grit levels prevents "confusion overload" and ensures incremental success, which builds confidence. Additionally, timely feedback loops that provide targeted guidance during struggles and celebrate milestones can reinforce persistence and motivation.

However, it is important to note that due to ethical considerations, the absence of a control group, the small sample size, and the short duration of the online collaborative peer feedback activity (only five weeks), this study does not allow us to establish a causal relationship between peer feedback and grit or generalize the findings to a broader population. The high probability of occurrence of two transitions (I-I and G-H) in both C1 and C2 groups also remains inconclusive regarding their relationship with grit.

Future research should extend the investigation of peer feedback mechanisms across diverse contexts, such as competitive versus cooperative settings, and over longitudinal timelines to validate causal relationships. By employing experimental or quasi-experimental designs with larger samples and longer durations, researchers can better understand the role of online collaborative peer feedback in fostering grit. Such studies may also explore whether shifts in online collaborative peer feedback behaviors and their effects on grit change in the context of intergroup competition. These efforts could provide deeper insights into how online education can become a robust platform for nurturing grit and essential non-cognitive competencies, ultimately supporting students in achieving long-term academic and personal goals.

REFERENCES

Adjei, S. A., Baker, R. S., & Bahel, V. (2021, June). Seven-year longitudinal implications of wheel spinning and productive persistence. In International Conference on Artificial Intelligence in Education (pp. 16–28). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-78292-4_2.

Alan, S., Boneva, T., & Ertac, S. (2019). Ever failed, try again, succeed better: Results from a randomized educational intervention on grit. The Quarterly Journal of Economics, 134(3), 1121–1162. https://doi.org/10.1093/qje/qjz006.

Alaoui, L., Fons-Rosen, C., (2016). Know when to Fold 'em: The Grit Factor. Working Papers 899, Barcelona Graduate School of Economics.

Almeda, Ma. V., & Asbell-Clarke, J. (2021). Scaffolding Executive Function in Game-Based Learning to Improve Productive Persistence and Computational Thinking in Neurodiverse Learners. HCI in Games: Serious and Immersive Games, 155–172. https://doi.org/10.1007/978-3-030-77414-1_12.

Aparicio, M., Bacao, F., & Oliveira, T. (2017). Grit in the path to e-learning success. Computers in Human Behavior, 66, 388–399. https://doi.org/10.1016/j.chb.2016.10.009.

Barber, W., van Oostveen, R., & Childs, E. (2019). Situating resilience, grit and growth mindset as constructs of social presence in the fully online learning community model (FOLC). Kidmore End: Academic Conferences International Limited. https://doi.org/10.34190/EEL.19.012.

Beck, J. E., & Gong, Y. (2013). Wheel-spinning: Students who fail to master a skill. In Artificial Intelligence in Education: 16th International Conference, AIED 2013,
Memphis, TN, USA, July 9-13, 2013. Proceedings 16 (pp. 431-440). Springer Berlin
Heidelberg. Botelho, A. F., Varatharaj, A., Patikorn, T., Doherty, D., Adjei, S. A., & Beck, J. E. (2019). Developing early detectors of student attrition and wheel spinning using deep learning. IEEE Transactions on Learning Technologies, 12(2), 158–170.

Buzzetto-Hollywood, N., Quinn, K., Wang, W., & Hill, A. (2019). Grit in online education. Journal of Education, Society and Behavioural Science, 30(4), 1–11.

Carbonneau, K. J. , Wong, R. M. , & Borysenko, N. . (2020). The influence of perceptually rich manipulatives and collaboration on mathematic problem- solving and perseverance. Contemporary Educational Psychology, 61(1), 101846.

Chen, Y. C., & Tsai, C. C. (2009). An educational research course facilitated by online peer assessment. Innovations in Education and Teaching International, 46(1), 105–117.

Cutts, Q., Cutts, E., Draper, S., O'Donnell, P., & Saffrey, P. (2010). Manipulating mindset to positively influence introductory programming performance. Proceedings of the 41st ACM technical symposium on computer science education, pp. 431–435.

Datu, J. A. D., Yuen, M., & Chen, G. (2018). The triarchic model of grit is linked to academic success and well-being among Filipino high school sstudents. School Psychology Quarterly, 33(3), 428–438.

Davies, P. (2000). Computerized peer assessment. Innovations in Education and Teaching International, 37(4), 346–355. https://doi.org/10.1080/135580000750052955

Dicerbo, K. E. (2014). Game-based assessment of persistence. Journal of Educational

Technology & Society, 17(1), 17–28.
Duckworth, A. (2016). Grit: The power of passion and perseverance. Scribner.
Duckworth, A. L., & Quinn, P. D. (2009). Development and validation of the short grit scale (grit-S). Journal of Personality Assessment, 91(2), 166–174. https://doi.org/10.1080/00223890802634290
Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. Journal of Personality and Social Psychology, 92(6), 1087–1101. https://doi.org/10.1037/0022-3514.92.6.1087
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.
Dweck, C., Walton, G. M., & Cohen, G. L. (2014). Academic tenacity: Mindset and skills that promote long-term learning. In Gates foundation. Bill & Melinda Gates
Foundation
http://web.stanford.edu/gwalton/home/Welcome_files/DweckWaltonCohen_2014.pdf
Farrington, C. A. (2012). Teaching Adolescents To Become Learners : The Role Of Noncognitive Factors In Shaping School Performance : A Critical Literature Review. University Of Chicago, Consortium On Chicago School Research.
Flores, R. M., & Rodrigo, M. M. T. (2020). Wheel-spinning models in a novice programming context. Journal of Educational Computing Research, 58(6), 1101–1120.
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. https://doi.org/10.1007/s11251-014-9319-4
Goegan, L. D., Le, L., & Daniels, L. M. (2023). Online Learning is a Rollercoaster: Postsecondary Students With Learning Disabilities Navigate the COVID-19 Pandemic. Learning Disability Quarterly, 46(3), 166-179. https://doi.org/10.1177/07319487221090912
Gray, J. P., and Mannahan, K. K. (2017). How well do trait measures of achievement predict students' perceptions of the link between personal effort and academic performance? J. Eff. Teach. 17, 16–27.
Hagger, M. S., & Hamilton, K. (2019). Grit and self-discipline as predictors of effort and academic attainment. British Journal of Educational Psychology, 89(2), 324–342
Hernández-Sellés, N., Galindo, J. M., Arteaga, O., & García, S. (2023). Diseño de un enfoque humanista con impacto social positivo en el ámbito tecnológico de las enseñanzas de grado: el caso del grado en diseño y gestión de proyectos transmedia. In P. C. Muñoz
Carril, C. Sarceda Gorgoso, E. J. Fuentes Abeledo & E. M. Barreira Cerqueiras (Eds.), La formación y la innovación educativa: ejes para la transformación social (141–164). Dykinson S.L. https://doi.org/10.2307/jj.2010047.11
Israel-Fishelson, R., & Hershkovitz, A. (2020). Persistence in a game-based learning environment: The case of elementary school students learning computational thinking. Journal of Educational Computing Research, 58(5), 891–918.
Kai, S., Heffernan, C., Almeda, M. V., Heffernan, N., & Baker, R. S. (2018). Decision tree modeling of wheel- spinning and productive persistence in skill builders. Journal of Educational Data Mining, 10(1), 36–71

Kench, D., Hazelhurst, S., & Otulaja, F. (2016). Grit and growth mindset among high school students in a computer programming project: A mixed methods study. In Annual Conference of the Southern African Computer Lecturers' Association (pp. 187–194). Cham: Springer

Kollar, I., Fischer, F., & Hesse, F. (2006). Collaboration scripts-A conceptual analysis. Collaboration Scripts-A Conceptual Analysis. Edu-Cational Psychology Review, 18(2), https://doi.org/10.1007/s10648-006-9007-2ï

Krumm, A. E., Beattie, R., Takahashi, S., Angelo, C. D., Feng, M., & Cheng, B. (2016). Practical measurement and productive persistence: Strategies for using digital learning system data to drive improvement. Journal of Learning Analytics, 3(2), 116–138.

Krumm, A., Coulson, A., & Neisler, J. (2022). Defining productive struggle in ST math: Implications for developing indicators of learning behaviors and strategies in digital learning environments. ACM international conference proceeding series, pp. 514–520

Lan, X., & Moscardino, U. (2019). Direct and interactive effects of perceived teacher-student relationship and grit on student wellbeing among staybehind early adolescents in urban China. Learning and Individual Differences, 69, 129–137. https://doi.org/10.1016/j.lindif.2018.12.003

Latifi, S., Noroozi, O., & Talaee, E. (2020). Worked example or scripting? Fostering students' online argumentative peer feedback, essay writing and learning. Interactive Learning Environments, 31(2), 655–669. https://doi.org/10.1080/10494820.2020.1799032

Latifi, S., Noroozi, O., Hatami, J., & Biemans, H. J. A. (2019). How does online peer feedback improve argumentative essay writing and learning? Innovations in Education and Teaching International, https://doi.org/10.1080/14703297.2019.1687005

Malureanu, A., Panisoara, G., and Lazar, I. (2021). The relationship between self-confidence, self-efficacy, grit, usefulness, and ease of use of eLearning platforms in corporate training during the COVID-19 Pandemic. Sustainability 13: 6633.

McGourty, J. (2000). Using multisource feedback in the classroom: A computer-based approach. IEEE Transactions on Education, 43(2), 120–124. https://doi.org/10.1109/13.848062

Mooradian, T.,Matzler, K., Uzelac, B.,& Bauer, F.(2016). Perspiration an inspiration: Grit and 1nnovativeness entrepreneuria success. .Journal o Economic Psychology,56.232–243.

Noroozi, O., Biemans, H., & Mulder, M. (2016). Relations between scripted online peer feedback processes and quality of written argumentative essay. The Internet and Higher Education, 31, 20–31. https://doi.org/10.1016/j.iheduc.2016.05.002

Nussbaum, M., Barahona, C., Rodriguez, F., Guentulle, V., Lopez, F., Vazquez-Uscanga, E., & Cabezas, V. (2021). Taking critical thinking, creativity and grit online. Educational Technology Research and Development, 69(1), 201–206. https://doi.org/10.1007/s11423-020-09867-1

Owen, V. E., Roy, M. H., Thai, K. P., Burnett, V., Jacobs, D., Keylor, E., & Baker, R. S. (2019). Detecting wheel-spinning and productive persistence in educational games. Proceedings of the 12th international conference on educational data mining (EDM 2019), pp. 378–38

Palaoag, T. D., Rodrigo, M. M. T., Andres, J. M. L., Andres, J. M. A. L., & Beck, J. E. (2016). Wheel-spinning in a game-based learning environment for physics. In

International Conference on Intelligent Tutoring Systems (pp. 234–239). Cham: Springer.

Peters, O., Koerndle, H., & Narciss, S. (2017). Effects of a formative assessment script on how vocational students generate formative feedback to a peer's or their own performance. European Journal of Psychology of Education, 1–27. https://doi.org/10.1007/s10212-017-0344-y

Porter, T., Molina, D. C., Blackwell, L., Roberts, S., Quirk, A., Duckworth, A. L., & Trzesniewski, K. (2020). Measuring mastery behaviours at scale: The persistence, effort, resilience, and challenge seeking (PERC) task. Journal of Learning Analytics, 7(1), 5–18.

Rodrigo, M. M. T., & Beck, J. B. (2014). Understanding wheel spinning in the context of affective factors. International Journal of People-Oriented Programming, 3(2), 75–90

Smart, K.L., & Cappel, J.J. (2006). "Students' Perceptions of Online Learning: A Comparative Study", Journal of Information Technology Education: Research, Vol. 5, No. 1, pp. 201–219. Retrieved from https://www.learntechlib.org/p/111541/

Strayhorn, T. L. (2014). What role does grit play in the academic success of black male collegians at predominantly white institutions? Journal of African American Studies, 18(1), 1–10. https://doi.org/10.1007/ s12111-012-9243-0

The Role of Classroom Contexts on Learners' Grit and Foreign Language Anxiety: Online vs. Traditional Learning Environment

Tough, P. (2012). How children succeed: Grit, curiosity, and the hidden power of character. Houghton Mifflin Harcourt.

Tsai, C. C. (2009). Internet-based peer assessment in high school settings. In Handbook of research on new media literacy at the K-12 level (pp. 743–754). IGI Global.

Tsai, C. Y. (2019). Improving students' understanding of basic programming concepts through visual programming language: The role of self-efficacy. Computers in Human Behavior, 95, 224–232.

Von Culin, K. R., Tsukayama, E., and Duckworth, A. L. (2014). Unpacking grit: motivational correlates of perseverance and passion for long-term goals. J. Posit. Psychol. 9, 306–312. doi: 10.1080/17439760.2014.898320

Wang, Y., Kai, S., & Baker, R. S. (2020). Early detection of wheel-spinning in assistments. In International Conference on Artificial Intelligence in Education (pp. 574–585). Cham: Springer.

Warshauer, H. K. (2015). Productive struggle in middle school mathematics classrooms. Journal of Mathematics Teacher Education, 18(4), 375–400.

Weegar, M.A. & Pacis, D. (2012). "A Comparison of Two Theories of Learning-Behaviorism and Constructivism as Applied to Face-to-Face and Online Learning", In Proceedings e-leader conference, Manila.

Wen, M. L., & Tsai, C. C. (2008). Online peer assessment in an inservice science and mathematics teacher education course. Teaching in Higher Education, 13(1), 55–67. https://doi.org/10.1080/13562510701794050

Yang, P. (2021). Exploring the Relationship Between Chinese EFL Students' Grit, Well-Being, and Classroom Enjoyment. Front. Psychol. 12:762945. doi: 10.3389/fpsyg.2021.762945

N. Law, G. Wong, E. Woo, & G. Jiang (Eds.) (2025). *Conference Proceedings of CITERS 2025.*
Hong Kong: The University of Hong Kong.

Transformative Synergy: The Innovative Practice of Generative AI and Digital Multimodal Composing in Language Education

Hong Cheng[1], Fengrui Ci[1], Rong Yu[1], Xiaochen Yu[1] and Lianjiang Jiang[1*]
[1]Faculty of Education, The University of Hong Kong, Hong Kong, China
[*] jljiang@hku.hk

With the rapid evolution of generative artificial intelligence (GenAI) and the increasing prevalence of multimodal communication, language education is undergoing unprecedented transformations. Traditional language pedagogy, which primarily focuses on linguistic skills, should shift towards a multimodal meaning-making approach to serve students' multiliteracies. Therefore, this symposium focuses on the deep integration of GenAI and digital multimodal composing (DMC), an approach that uses digital tools to create compositions combining text, sound, images, video, and other semiotic resources. Presenters aim to explore its transformative impact on language learners and teachers through empirical research on teaching practices. By presenting four empirical studies, the symposium will reveal how GenAI empowers DMC in language learning and teaching while addressing the potential constraints and challenges learners and teachers face in the AI era.

First, Yu Rong will present a qualitative study involving five Chinese undergraduate students, inquiring how they use GenAI to assist in L2 writing in a DMC project. Through screen recordings, reflective journals, and interview analyses, the study identifies three patterns of GenAI usage: linguistic expression, content analysis, and idea generation, while GenAI translation plays a significant role in L2 writing. The findings suggest that the linguistic mode remains dominant in L2 writing in DMC, and students' limited linguistic competence leads to increased reliance on AI, reinforcing the monomodal nature of the writing process. Consequently, the study underscores the need for educators to support students in developing linguistic proficiency while cultivating multimodal awareness through GenAI tools.

In the second presentation, Yu Xiaochen will explore how GenAI-assisted DMC fosters multilingual students’ creativity in a secondary ESL classroom in Hong Kong. Based on the analysis of video recordings, student interviews and digital multimodal artifacts, the study reveals that GenAI-assisted DMC empowers students to create a translanguaging space, where they creatively express their multicultural identities by freely drawing on their funds of knowledge, multilingual resources, and integrating various semiotic modes. The research advocates that GenAI-assisted DMC facilitates the creation of translanguaging spaces, which not only foster students' creativity in self-expression but also develops their multilingual and multicultural identities.

In the third presentation, Cheng Hong will explore the potential role of GenAI in supporting teachers in providing feedback on students’ DMC (e.g., posters). Grounded in the theoretical foundations of social semiotics and multiliteracies, her study compares the types and foci of teacher feedback with ChatGPT-generated feedback by L2 English teachers at a Chinese university. It also explores their perception of integrating ChatGPT into DMC feedback practices through semi-structured interviews and reflective journals. The study finds that ChatGPT-generated feedback helps teachers expand their focus from linguistic elements to encompass the orchestration of multiple modes (e.g., textual, visual, spatial) and reflect on students' choices to maximize the affordance of each semiotic mode. However, teachers require further guidance in understanding intermodal relationships that contribute to multimodal creation as a coherent ensemble. These findings suggest that ChatGPT can help foster semiotic awareness of teachers to develop a more comprehensive and critical view of DMC and provide effective feedback on students' multimodal meaning-making practices in a digital age.

Finally, Ci Fengrui will present a collaborative case study investigating how pre-service language teachers construct their role identity in GenAI-assisted DMC tasks. Focusing on pre-service and in-service teachers enrolled in an MA TESOL programme in Hong Kong, the study employs the Dynamic Systems Model of Role Identity (DSMRI) to analyse how teachers' professional identities evolve from being "instructors of cognitive language skills" or "determiners of learning resources. " to becoming "facilitators of multimodal integration" or "reflectors of composition" when using GenAI to complete DMC tasks. Exploring how teachers develop their role identity to engage with GenAI and DMC provides practical insights into why and how AI cannot replace teachers. Through the lens of GenAI-assisted DMC, this study provides valuable implications for designing language teacher training programmes that emphasise digital, multimodal, and AI-integrated teaching contexts.

To sum up, this symposium aims to demonstrate that GenAI is neither a substitute of language teachers nor a threat to the necessity of language learning. Instead, it serves as a vital tool for fostering pedagogical transformation and innovation in language education. Through GenAI-assisted DMC, teachers can facilitate students to transcend linguistic and modal limitations, enabling them to strengthen their language knowledge while achieving deeper meaning-making and identity expression in multimodal composing. Integrated with DMC, GenAI is a catalyst for teachers' professional development and identity expansion, encouraging them to actively learn, embrace, and reflect on GenAI and other technologies. This symposium provides a solid foundation and direction for exploring GenAI-DMC as a language pedagogy and other AI-assisted teaching practice.

Keywords: Generative AI, Digital Multimodal Composing, L2 Education, Language Pedagogy and Practice

N. Law, G. Wong, E. Woo, & G. Jiang (Eds.) (2025). *Conference Proceedings of CITERS 2025.*
Hong Kong: The University of Hong Kong.

Navigating Information Retrieval in the Age of AI: Challenges, Strategies, and Practical Insights

Muhammad ALI[1*] & Gary K.W. WONG[1]
[1]Faculty of Education, The University of Hong Kong, Hong Kong, China
[*] akula@connect.hku.hk

Abstract: As artificial intelligence (AI) transforms information retrieval, traditional search models are evolving, introducing both opportunities and challenges. This interactive workshop explores how AI-powered agents reshape retrieval processes, influencing accuracy, reliability, and transparency. Through comparative analysis, case studies, and hands-on exercises, participants will explore AI-driven ranking, query interpretation, and prompt engineering. Designed for researchers and professionals, whether experts or novices, this session equips attendees with practical strategies to optimize AI-assisted search and retrieval.
Keywords: information retrieval, prompt engineering, algorithmic bias

In an era where artificial intelligence (AI) is redefining digital interactions, information retrieval is becoming increasingly complex. This interactive workshop explores how AI-powered agents are transforming traditional search and retrieval systems, creating new opportunities while introducing critical challenges related to bias, accuracy, reliability, and transparency.

The session begins with a comparative analysis of conventional information retrieval models and AI-enhanced systems, highlighting key differences in how information is processed, ranked, and presented. While AI agents simulate human-like reasoning, they introduce both efficiency and unpredictability, raising concerns about algorithmic bias, misinformation, and contextual misinterpretation.

Through interactive discussions, case studies, and hands-on exercises, participants will explore the mechanics of AI-driven retrieval, gaining insights into how AI agents interpret queries, structure responses, and influence information consumption. A key focus will be on prompting techniques and prompt engineering, equipping participants with practical strategies to refine AI interactions for more precise, reliable, and context-aware information retrieval.

Designed for researchers, professionals, and AI enthusiasts, this workshop offers comprehensive, applied learning experience, empowering attendees to pragmatically navigate AI-driven retrieval systems and develop effective methodologies for optimizing AI-assisted search processes.

N. Law, G. Wong, E. Woo, & G. Jiang (Eds.) (2025). *Conference Proceedings of CITERS 2025.*
Hong Kong: The University of Hong Kong.

Enhancing AI Literacy in Higher Education: Library-Led Training and Resources for Student Empowerment

Alice CHAN[1], Cindy LIANG[1], Rebecca TAM[1]
[1] The University of Hong Kong, Hong Kong, China
* mailto:alicectc@hku.hk

The objective of this workshop is to share AI literacy training initiatives and resources designed by the University of Hong Kong Libraries to enhance students' understanding of AI concepts and its implications. We will present our tailored AI Literacy guide, which provides foundational knowledge on AI, its applications across disciplines, and the ethical considerations surrounding its use at a tertiary level.

The session will focus on sharing comprehensive AI seminar series and workshops hosted by the University Libraries, specifically hands-on training sessions with a focus on practical applications of AI tools. Participants will learn about instructional strategies for creating educational resources that empower students to engage with AI technologies critically.

To foster rich conversations, participants will be encouraged to share their experiences and challenges in implementing AI literacy initiatives at their institutions. This collaborative approach will facilitate the exchange of ideas and best practices, creating a supportive environment for professional growth.

The academic significance of this workshop lies in its potential to promote inclusive and effective educational practices through AI literacy. By equipping educators and librarians with insights into successful initiatives, we aim to enhance the overall learning experience for students in a technology-driven world.

N. Law, G. Wong, E. Woo, & G. Jiang (Eds.) (2025). *Conference Proceedings of CITERS 2025.*
Hong Kong: The University of Hong Kong.

Promoting Pedagogical Innovation in STEAM Education at Secondary and Primary Schools Through External and Internal Networks

Shihui FENG[1*], Jiaqi Xu[1], Siyou WU[1]
[1] The University of Hong Kong, Hong Kong, China
[*] shihuife@hku.hk

Abstract: This professional workshop focuses on sharing research findings and practical experience on leveraging external and internal networks to prompt pedagogical innovation in primary and secondary STEAM education in Hong Kong. Featuring three talks by the project team at HKU and experienced STEAM coordinators at local schools. Through theoretical insights and practical cases, this workshop will equip participants with strategies for integrating external resources and internal expertise in STEAM education, promoting a culture of collaboration.
Keywords: STEAM education, STEAM networks, pedagogical innovation

This one-hour workshop is dedicated to share research findings and practical experience regarding how external and internal networks can be employed to innovative pedagogical approaches to advance secondary-level and primary-level STEAM education in Hong Kong. The workshop will include three talks provided by the HKU project team led by Professor Shihui Feng and two experienced STEAM coordinators from local schools. The HKU project team will share their research findings on the impact of external collaborators and internal stakeholders on STEAM pedagogical innovation in secondary and primary schools, and STEAM coordinators will share practical experience on leveraging these networks in promoting STEAM education particularly with emerging digital technologies. Through a combination of theoretical insights and practical cases, participants will gain a deeper understanding of the critical role that external and internal networks as well as learn creative strategies for integrating external resources and internal expertise to foster pedagogy in STEAM education. Attendees will be equipped with practical insights and strategies to effectively leverage external and internal networks to promote pedagogical innovation in STEAM education, ultimately promoting a culture of creativity and collaboration in teaching and learning STEAM in this digital world.

Technology-Supported Learning Design Process with Design-Aware Learning Analytics

Hongfeng LIU[1], Wai Hung LAM[1], Jiewen FENG, Pakon KO[1]*
[1] Centre for Information Technology in Education, Faculty of Education,
The University of Hong Kong, Hong Kong, China
* kopakon@hku.hk

Abstract: Transitioning from traditional content-driven teaching to student-centered learning poses design challenges for teachers. Often, key learning design principles may be overlooked during curriculum development. This interactive workshop demonstrates how the IDEALS system, with embedded design principles and a thinking process informed by the Learning Design Triangle framework, supports teachers in designing meaningful learning experiences for students and in using learning analytics to enhance teaching and learning.
Keywords: learning design, learning analytics, learning experience, student-centered learning, learning design principles, IDEALS system

In the era of rapid technological and social changes, it is crucial for students to acquire essential future skills, such as collaborative problem-solving skills and digital competency (OECD, 2020). A student-centered learning approach is advocated because it enables students to engage actively in inquiry-based learning and problem-solving, thereby developing their competencies.

The shift from a traditional content-driven teaching to student-centered learning presents a learning design challenge to teachers. To address this, the IDEALS system has been developed. It aims to provide both conceptual and technological scaffolds to assist teachers in designing a meaningful learning process for students.

The learning design process in the IDEALS system is informed by the Learning Design Triangle (LDT) framework developed by Law and Liang (2020). It highlights learning design decisions such as intended learning outcomes, disciplinary practice, pedagogical approach, and the overall learning process in terms of the sequence of curriculum components. The LDT framework echoes different levels of pedagogical decision-making (Goodyear, 2005) in the learning design process. This design framework ensures the inclusion of learning analytics questions and solutions within the decision-making process.

The IDEALS research is methodologically informed by Design-Based Implementation Research (Fishman & Penuel, 2018) with Research Practice Partnership (Coburn & Penuel, 2016). We engage in-service teachers in using the IDEALS system to design and implement curriculum units. Teachers' perspectives have been collected through interviews, usability tests and teachers' sharing sessions for evaluating the system.

The interactive workshop will feature a live demonstration of the IDEALS system, complemented by hands-on exploration of its chatbot and design resources (pattern libraries). Through hands-on activities, participants can explore the learning design thinking process, as supported by the systems, from the perspective of actual users. By sharing authentic school cases and implementation examples, participants will gain practical strategies for designing effective, student-centered curriculums across various disciplines.

ACKNOLWDGEMENT

Intelligent DEsign-Aware Learning analytics empowered 21C Learning & Teaching System

(IDEALS) project is funded by the Quality Education Fund, the e-Learning Ancillary Facilities Programme (Project No. 2021/0212) and the Research Grants Council of the HKSAR Government, General Research Fund (Grant No. 17610423).

REFERENCES

Coburn, C. E., & Penuel, W. R. (2016). Research–practice partnerships in education: Outcomes, dynamics, and open questions. Educational researcher, 45(1), 48-54.

Fishman, B., & Penuel, W. (2018). Design-based implementation research. In International handbook of the learning sciences (pp. 393-400). Routledge.

Goodyear, P. (2005). Educational design and networked learning: Patterns, pattern languages and design practice. Australasian journal of educational technology, 21(1).

Law, N., & Liang, L. (2020). A multilevel framework and method for learning analytics integrated learning design. Journal of Learning Analytics, 7(3), 98-117.

OECD. (2020). What Students Learn Matters Towards a 21st Century Curriculum. OECD Publishing.

Fostering Resilient Educational Systems Through STEAM Innovation Networks

Shihui FENG[1*], Jiaqi XU[1], Siyou WU[1]
[1] The University of Hong Kong, Hong Kong, China
[*] shihuife@hku.hk

Abstract: This panel discussion will explore collaborative networks in enhancing STEAM education sustainability in Hong Kong. Panelists will analyze how external collaboration networks can foster connections among various stakeholders, address challenges of resource constraints and limited professional development, and ultimately provide policy suggestions for promoting effective STEAM initiatives. Attendees will gain new insights into external collaboration in STEAM education and practical recommendations for advancing STEAM education at a systemic level.
Keywords: STEAM education, STEAM networks, external collaboration

This one-hour panel discussion will explore the significant role of collaborative networks in enhancing STEAM (Science, Technology, Engineering, Arts, and Mathematics) education sustainability as well as the policy suggestions for fostering networks and connections among various stakeholders for STEAM education in Hong Kong. The session will be chaired by Prof Shihui Feng from Faculty of Education, University of Hong Kong, and the panelist will consist of five members, representing the key stakeholder types identified from the project, including primary and secondary school principals, STEAM coordinators, association leaders, and academic stakeholders. The panel discussion will discuss the impacts of external and internal networks on STEAM education, the mechanisms for developing STEAM school networks, as well as the relationships between emerging digital technologies and school STEAM networks. The panelists aim to offer practical insights into how these connections can be leveraged to promote sustainable and impactful STEAM initiatives within educational systems. This panel discussion will not only enrich attendees' theoretical understanding of network effects on STEAM education but also provide practical recommendations for policymakers aiming to advance STEAM education at a systemic level.

N. Law, G. Wong, E. Woo, & G. Jiang (Eds.) (2025). *Conference Proceedings of CITERS 2025.*
Hong Kong: The University of Hong Kong.

VR Creation Experience in Environmental Education for Primary School Students: A Preliminary Exploration

Ka Wai LAU[1*], Xiao HU[1]
[1] The University of Hong Kong, Hong Kong, China
[*] kawailau@connect.hku.hk

Abstract: While virtual reality (VR) is increasingly used in education, few studies have focused on VR creation in environmental education. This study involved 12 primary schools, with the goal of examining students' experiences in creating VR content about environmental issues. Field trips and VR content creation workshops were held for 85 students. Preliminary findings based on survey and interview responses illustrate that VR content creation enhanced students' environmental awareness and helped them acquire digital technology skills.
Keywords: digital literacy, virtual reality creation, primary education, digital maker activities, environmental education

INTRODUCTION

In this study, our research team leverages a low-tech VR content creation approach, to teach primary school students to create VR stories in the context of an environment conservation project [1]. The participants were primary school students from the fourth to sixth grades, who utilized tablets and an online VR creation platform to produce VR stories related to the ecosystem. Ninety-one (91) students from 12 schools participated in this study, with valid survey responses obtained from 85 students. Through the analysis of these responses, the study seeks to address the following research question:

RQ1: What are the perceptions and experiences of primary school students regarding the use of VR creation in enhancing their environmental awareness and digital technology skills?

The research question seeks to explore students' perspectives on VR content creation in the context of environmental education. The findings will provide empirical insights regarding the benefits and obstacles of incorporating VR content creation in environmental education, providing valuable guidance for pedagogical design and the educational use of VR creation tools.

THE VR CREATION PROGRAM

In the VR content creation program, students from 12 primary schools in Hong Kong, aged 8 to 12, used an online VR content creation platform to develop stories about local ecological sites. They took panoramic photos during field trips, researched species, and created narrative scripts that introduce their VR stories. The program aimed to raise awareness of environmental issues and showcase local ecosystems using digital tools. Two workshops were held in each participating school to guide students in researching species, creating scripts, and using the VR content creation platform (Figure 1) named Collaborative Learning Environment for VR creation (CLEVR). This platform is designed to be easy to use, allowing users to create VR stories with text or audio descriptions. It is specifically designed to support students of various levels of technological background in engaging in maker activities [3]. In the CLEVR platform, uploaded photos, added annotations and audio, sequenced scenes, narrations as well as added navigation buttons between scenes to create immersive VR stories[2].

Figure 1: the interface of the CLEVR system, the editor view with panorama content in the middle

PRELIMINARY RESULTS

The participants filled a pre-questionnaire before joining the VR story creation program and a post-questionnaire immediately after submitting the finalized VR stories. Both questionnaires measured their self-reported digital literacy. Item were rated by a 7-point Likert scale from 1 = I strongly disagree to 7 = I strongly agree. Table 1 shows the average responses to each question. A paired t-test was conducted to compare students' responses to each of the questions, and the results are also reported in Table 1.

Table 1. Participants' digital literacy measured before and after the VR content creation program

Questions	Pre-	Post-	p-value
1. I keep up with the latest digital technology	5.02	5.54	0.01*
2. I can quickly learn new digital technologies	5.11	5.55	0.03*
3. I can design and produce digital media.	4.49	5.44	0.00*
4. I can present digital media to the public	4.78	5.43	0.14
5. I know how to solve technical problems in digital technology.	4.88	5.40	0.03*

*: significant at $p < 0.05$ level.

The findings indicate that students were largely very positive about their VR content creation experience. After the field trip, on average, 75% of respondents reported an increase in their knowledge of nature conservation. Additionally, 86% of respondents agreed that VR virtual scenes could help people learn about the things presented in them.

After the workshops, 87% of respondents agreed that digital technology enhanced their awareness of community environment and nature conservation. 71% of respondents strongly agreed or agreed that "VR content production workshop (students)" increased their interest in other environmental activities. Additionally, 67% and 64% of participants respectively strongly agreed or agreed that this workshop helped them practice environmental protection in their daily lives and promote environmental conservation to family, classmates, and friends, dedicating efforts to support nature conservation. Last but not least, 77% of respondents reported that they were confident to use the CLEVR platform and create the VR content on their own.

4. CONCLUSION AND FUTURE WORKS

This study leverages a low-tech VR content creation approach to teach primary school students to create VR stories for the purposes of promoting environment conservation. The study found that students were positive about creating VR content, regarding it as a valuable opportunity to develop technology skills and improve environmental awareness. The simple process of using a tablet for VR creation was beneficial, especially for students with limited technology exposure. Workshops providing guidance and support were crucial for students to gain confidence in using digital tools. The low-tech barrier approach to VR creation was effective in teaching information technology in primary education. Students enjoyed collaborating with peers and deepened their understanding of ecological issues through VR content creation. By creating VR stories, students transition from being passive consumers to active creators of online resource platforms. This approach also promotes students' learning initiative and enhances their creativity. There are some limitations of the current study. First, due to restrictions on sample size and resulting power, this study focused on the underlying mechanism following the digital literacy perspective on VR content creation by pre-test and post-test questionnaire survey. Additional unexplored areas of this study include: 1) the technological barrier, including access to VR devices and digital literacy levels, and 2) student engagement differences based on age or prior tech experience. As an ongoing study, interviews conducted with students and teachers are being analyzed to complement the preliminary results. Future work will also include assessing the issues and challenges encountered during the VR content creation process in primary education.

ACKNOWLEDGEMENT

This research project (Project Number: EE&CA2946) is partially funded by The Environment and Conservation Fund of The Government of the Hong Kong Special Administrative Region.

REFERENCES

[1] Gestiada, R. J., Tisoy, F. J., & Lasala, N. J. (2025). The 360˚ View: Contextualized Virtual Reality Tours as Innovative Teaching Tool in Ecology for Elementary School Students. Journal of Basic Education Research, 6(1), 23–36. https://doi.org/10.37251/jber.v6i1.1213

[2] Hu, X., Ng, J., & Lee, J. H. (2019). VR creation experience in cultural heritage education: A preliminary exploration. Proceedings of the ASIST Annual Meeting, 56(1), 422–426. https://doi.org/10.1002/pra2.42

[3] Southgate, E., Grant, S., Ostrowski, S., Norwood, A., Williams, M., & Tafazoli, D. (2022, March). School students creating a virtual reality learning resource for children. In 2022 IEEE

conference on virtual reality and 3D user interfaces abstracts and workshops (VRW) (pp. 01-06). IEEE. https://doi.org/10.1109/VRW55335.2022.00060

N. Law, G. Wong, E. Woo, & G. Jiang (Eds.) (2025). *Conference Proceedings of CITERS 2025.*
Hong Kong: The University of Hong Kong.

EFL University Students' Emotions and Cognitive Loads in Chatbot-Supported Versus Peer-Supported Reading Activities: A Comparative Study

Yuanke LI[1*], Xiaotian GUAN[2]
[1] School of Foreign Studies, South China Normal University, China
[2] Faculty of Education, The University of Hong Kong, Hong Kong, China
[*] liyuanke@m.scnu.edu.cn

Abstract: Using a within-subject design, this study engaged 60 Chinese EFL university students in reading English academic passages either with Gen AI chatbot support or with peer support. We compared the levels of positive and negative emotions and intrinsic and extraneous cognitive loads reported by students in these two reading conditions and conducted student interviews to delve into the factors affecting their emotions and cognitive loads in the two interactive reading conditions.
Keywords: human-chatbot interaction, peer interaction, emotion, cognitive loads, academic reading

1. RESEARCH OBJECTIVES

This study aims to compare EFL university students' emotions and cognitive load (CL) during chatbot-supported English academic reading to peer-supported reading. The research questions (RQs) for this study are:

1. Are there any significant differences in students' emotions and cognitive load when reading with a chatbot compared to reading with peer?

2. What are the factors contributing to the differential effects on EFL students' emotions and cognitive load, if any, between the two interactive reading conditions?

2. CONCEPTUAL FRAMEWORK

Pekrun's (2006) Achievement Emotion Theory categorizes learners' emotion into four types: positive-activating emotion (PAE), positive-deactivating emotion (PDE), negative-activating emotion (NAE), and negative-deactivating emotion (NDE).

Sweller et al.'s (2019) model of cognitive load (CL) underscores three aspects: intrinsic cognitive load (ICL), extraneous cognitive load (ECL), and germane processing.

3. METHODOLOGY

60 English major students from two parallel classes in a Chinese university participated in this study.

A within-subjects counter-balanced design was employed to examine potential variations in the levels of emotions and CL experienced by students in two peer-supported reading activities and two chatbot-supported reading activities during five weeks.

Post-activity questionnaire survey was conducted at the end of each reading activity to evaluate the levels of their emotions and CL. Additionally, 16 students voluntarily participated in the semi-structured interviews after each of the two rounds of experiment, during which the influential factors of their emotion and CL were explored.

To address RQ 1, the non-parametric Wilcoxon signed-rank tests were used to examine the differences in students' emotions and CL. Regarding RQ 2, interview responses were analyzed using thematic analysis (Braun & Clarke, 2006).

4. RESULTS

Concerning RQ1, students experienced a significantly higher level of PAE ($z = 2.941$; $p = 0.003$; $r = 0.404$) and PDE ($z = 2.846$; $p = 0.004$; $r = 0.391$), but a significantly lower level of NAE ($z = 3.474$; $p = 0.001$; $r = 0.477$) when reading with chatbot. However, students experienced a similar level of NDE in the two conditions ($p=0.216$). Additionally, students experienced a significantly lower level of ICL ($z = 2.253$; $p = 0.024$; $r = 0.309$) and ECL ($z = 2.827$; $p = 0.005$; $r = 0.388$) when reading with chatbot.

Regarding RQ2, thematic analysis found that reading with AI chatbot demonstrates superior effects in cultivating positive emotion through high quality of support and low-stress learning environment while showing limitations in emotional exchange.

In terms of CL, the different quality of the collaborator's support is identified to contribute to the differences in ICL under the two conditions. Additionally, reading with AI chatbot showed mixed effects regarding CL management: the convenience of tracking discussion records reduces students' ECL, while the complex responses of the chatbot and the complicated communication repair with the chatbot may trigger additional CL.

5. CONCLUSION

Overall, this study has elucidated the mechanisms underlying differential effects of reading with AI chatbot support versus peer support on students' emotions and CL.

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

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. https://doi.org/10.1007/s10648-006-9029-9

Sweller, J., Van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive Architecture and Instructional Design: 20 Years Later. *Educational Psychology Review, 31*(2), 261–292. https://doi.org/10.1007/s10648-019-09465-5

N. Law, G. Wong, E. Woo, & G. Jiang (Eds.) (2025). *Conference Proceedings of CITERS 2025.*
Hong Kong: The University of Hong Kong.

Creating Intelligent Learning Environments for Equitable Access and Digital Wellbeing in Math Education

Wenqi LIU[1*]
[1] Pengzhi Education
[*] mailto:vickiliu@alumni.upenn.edu

Abstract: The application of digital technologies in the field of education has brought opportunities to enhance equitable access and promote digital wellbeing. This proposal explores how Banbuniuniu, an online math learning platform that we designed can help to create an intelligent learning environment. This study will evaluate the effectiveness of the platform in promoting education equity and digital wellbeing in math education.
Keywords: digital wellbeing, education equity, math learning, intelligent learning environment

1. OBJECTIVES AND PURPOSES

This study aims to evaluate and explore how different features, design, and modules that the online math learning platform Banbuniuniu can help to positively impact students learning and promote education equity as well as digital wellbeing through providing users with an easily accessible, inclusive and engaging learning environment.

2. THEORETICAL FREAMWORK

There are 17 percent of children, youth, and adolescents worldwide out of school. The ratio is much higher in developing countries (Schmelkes, 2020). Achieving education equity is a complex challenge due to social, economic, and educational factors. Quality education as one of the United Nations' sustainable development goals (SDGs) specifically aims to "ensure inclusive and equitable quality and promote lifelong learning opportunities for all" (United Nations, 2015). The study will analyze how Banbunniuniu can help with this goal.

This study draws on the new framework for assessing the impact of technology on wellbeing, which provides a comprehensive method to measure digital welling in the aspects of physical health, mental health, and social connections (Jisc, 2020). Besides, prior research related to digital wellbeing will be incorporated in this study, which emphasizes the importance of digital tools and online learning on students' academic performance and engagement (Selwyn et al., 2020).

3. METHODS AND EXPECTED RESULTS

To evaluate the impact of Banbuniuniu on digital wellbeing and education equity, the study will apply a Randomized Controlled Trial (RCT) Design as the research method. A total of 1,000 students from diverse socio-economic backgrounds are randomly assigned into two groups and each group has 500 students. Students in the treatment group will use Banbuniuniu for personalized math learning. Students in the control group have no access to Banbuniuniu and will follow traditional math curriculum. They will be continuously following the assigned learning instructions for 12 months. During the process, quantitative data such as students' performance scores and learning time will be collected. Qualitative data like the questionnaires from teachers and parents, interviews from students about the Banbuniuniu's usability and accessibility will be collected. Both quantitative and qualitative data will be analyzed to measure how Banbuniuniu can help to improve education equity and digital wellbeing.

We hypothesize that students in the treatment group have a better academic performance, engagement, and improved learning focus than students in the control group. It indicates that

Banbuniuniu is effective to support students' learning outcomes. We also expect that students in the treatment group have better learning experience, less tech-related stress, and greater feeling of engagement and support because of the diverse and intelligent learning environment. In terms of equity, students in the treatment group can easily access to various modules such as daily challenges, math theater, math comics, and math battles. They are able to be exposed to an engaging, fun and immersive learning environment with less anxiety and more confidence. They can also get connected with their peers and AI tutors through this platform' s community features, which narrowing the learning gaps among students from different socioeconomic backgrounds. By contrast, students who learn math using traditional methods, may not experience these benefits to the same extent. If the data support the expected outcomes, it will suggest that an intelligent learning environment like Banbuniuniu can play an essential role in improving digital wellbeing and equitable access to math education.

4. CONCLUSION

In summary, this research will illustrate the potential of intelligent learning environments to promote education equity and enhance learners' digital well-being. The results of this study are expected to have both academic and practical significance for learners, educators, policymakers, and tech designers. If Banbuniuniu is demonstrated to be effective and practical, it will offer a potential to future math teaching and learning. It can either to be leveraged as a supplement of a traditional classroom, or it can be offered to students who have limited educational resources, getting them access to quality education and intelligent learning experience.

REFERENCES

Jisc. (2020). Digital wellbeing: A new framework for assessing the impact of technology on wellbeing. Jisc. https://www.jisc.ac.uk

Schmelkes, S. (2020). Recognizing and overcoming inequity in education. UN Chronicle. Retrieved from https://www.un.org/en/un-chronicle/recognizing-and-overcominginequity-education

Selwyn, N., Hillman, T., Eynon, R., Ferreira, G., Knox, J., Macgilchrist, F., & Sancho-Gil, J. M. (2020). What's next for Ed-Tech? Critical hopes and concerns for the 2020s. *Learning, Media and Technology, 45*(1), 1–6. https://doi.org/10.1080/17439884.2020.1694945

United Nations. (2015.). Goal 4: Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all. Sustainable Development Goals. Retrieved from https://sdgs.un.org/goals/goal4

N. Law, G. Wong, E. Woo, & G. Jiang (Eds.) (2025). *Conference Proceedings of CITERS 2025.*
Hong Kong: The University of Hong Kong.

Becoming Learning Designers: Expert Guidance and AI Collaboration in a Five-Iteration Case Study of K-12 STEM Curriculum Redesign

Liu Yong[1*], Liu Kun[2], Pan Pengfei[3], Chen Jun[1], Luo Guilin[4]
[1] Hongfan School Attached to Chongqing No. 8 Secondary School University, Hong Kong, China
[2] The University of Hong Kong, Hong Kong, China
[3] Graduate School of Education, Peking University, China
[4] Chongqing Central Park Primary School, China
[*] liukun2014@gmail.com

Abstract: Teachers face significant challenges transitioning from teacher-centered to inquiry-based learning, particularly in reconceptualizing pedagogy and design assumptions. This study explores how expert conceptual scaffolding with GenAI's practical support facilitates this shift. Through five co-design cycles of a 'Balancing Bird' lesson with an experienced secondary STEM teacher, we trace a four-phase developmental trajectory toward reflective innovation. Findings demonstrate how this synergy enables pedagogical reflection and sustainable redesign, offering a model for teachers' shift into learning designers.
Keywords: teacher learning, learning design, human-AI collaboration, STEM

This study investigates how a secondary school teacher transitioned into a learning designer through five iterative cycles of the Balance Bird Production curriculum redesign, supported by expert mentorship and generative AI (ChatGPT).

Grounded in theories of Laurillard's Teachers as a Learning Designers, learning design (Mor et.al, 2015) and Human AI Collaboration (Lee et.al, 2025), the research employs a longitudinal case study analyzing teacher reflection journals, expert feedback logs, and AI interactions through thematic coding aligned with the Dreyfus model of skill acquisition. Findings reveal a four-phase transformation: (1) Novice Dependence, leveraging expert frameworks (e.g., 7-step scientific inquiry) and AI-generated real-world examples; (2) Guided Experimentation, integrating AI suggested hypothesis-driven tasks with expert-decomposed inquiry elements; (3) Strategic Integration, balancing AI-enhanced failure analysis (e.g., "Why might the bird tip?") and expert rubrics for peer assessment; (4) Reflective Innovation, contextualizing designs via student-led design challenges.

Findings reveal teachers' transformation from procedural instruction to inquiry-based design, driven by complementary inputs: experts provided pedagogical scaffolding, while AI offered granular creative solutions (e.g., question chains, quantitative metrics), requiring teacher-led critical filtering. Thematic analysis also demonstrated the teacher's growing agency in prioritizing inputs—rejecting misaligned AI suggestions (e.g., premature quantitative analysis) while adopting expert-guided variable control strategies. Student outcomes improved significantly in structured scientific reasoning and creative prototypes across iterations.

The study proposes a scalable "Expert-AI-Teacher" triad model, balancing AI's democratizing potential (e.g., rapid ideation) with expert coherence in pedagogical design. Implications advocate for professional development programs that blend AI tools with expert communities of practice, empowering teachers as adaptive, critical designers in STEM education.

REFERENCES

Laurillard, D. (2013). Teaching as a design science: Building pedagogical patterns for learning and technology. Routledge.

Mor, Y., Ferguson, R., & Wasson, B. (2015). Learning design, teacher inquiry into student learning and learning analytics: A call for action. *British Journal of Educational Technology, 46*(2), 221- 229.

Lee, S. Y., Law, M., & Hoffman, G. (2025). When and how to use AI in the design process? Implications for human-AI design collaboration. *International Journal of Human – Computer Interaction*, *41*(2), 1569-1584.

N. Law, G. Wong, E. Woo, & G. Jiang (Eds.) (2025). *Conference Proceedings of CITERS 2025.*
Hong Kong: The University of Hong Kong.

Applying the Learning Design Studio in Interprofessional Practice Between Speech-Language Pathologists and Classroom Teachers

Eddy C. H. WONG[1*], Jeremy Tzi Dong NG[1]
[1] Faculty of Education, The University of Hong Kong, Hong Kong, China
[*] eddychw@hku.hk

Abstract: This presentation discusses the Learning Design Studio (LDS), an online platform to enhance interprofessional practice between speech-language pathologists (SLPs) and teachers. By providing a framework for collaborative course design, the LDS aims to improve speech and language skills in students with and without communication disorders. The presentation will include a hypothetical design for students with communication disorder and outline future research on implementing this collaborative approach in schools.
Keywords: learning design, interprofessional collaboration, speech-language pathology, communication disorders

1. INTRODUCTION

Interprofessional practice between speech-language pathologists (SLPs) and teachers is one of the scopes in local school-based speech therapy services (Education Bureau, 2023). It aims to improve and enhance speech and language abilities of students with and without communication disorders. However, there are insufficient evidence-based guidelines to support this cross-profession collaborative service (Armstrong et al., 2023; Mathers et al., 2024), which can improve practices and also offer professional development opportunities for both SLPs and teachers. This presentation aims to discuss the use of the Learning Design Studio (LDS), a theoretically driven online platform for supporting learning design, to facilitate collaborative SLP-teacher classroom-based teaching, and most importantly, supporting children with speech and language needs in the classroom.

Grounded in multi-level learning design frameworks (Law & Liang, 2020), the LDS is a tool that aims to facilitate design thinking and provide strategies and tools to foster and support learning designs, empowering teachers and other professionals in education settings as learning designers (Law et al., 2017). In essence, the LDS emphasizes three levels of design, including (1) course level, (2) learning unit level, and (3) learning task level, encompassing various elements ranging from intended learning outcomes to sequences of learning tasks. A major objective of the LDS is to help learning designers maintain alignment between these elements on different levels, ultimately aiming to facilitate students' learning.

It is expected that the application of LDS in the SLP-teacher interprofessional practice will offer both a robust framework and an online platform for SLPs and teachers to co-designing courses or lessons aimed at improving and enhancing students' speech and language skills, regardless of whether they have communication disorders. In this presentation, we will discuss a hypothetical course design focused on enhancing storytelling skills in primary three students.

2. A HYPOTHETICAL CASE

Our approach incorporates perspectives from both teachers and SLPs to create an effective learning experience. Teachers bring a strong understanding of the curriculum, particularly regarding vocabulary use and complex sentence structures in Chinese Language, while SLPs contribute flexible learning strategies catered to children with communication disorders. Storytelling skills are targeted because these skills are assessed at various stages of school-age development through the Territory-wide System Assessment, providing schools with valuable data on students' language performance.

The course consists of four weekly one-hour lessons. Grounded in the multi-level learning design framework proposed by Law and Liang (2020), we have established clear intended learning outcomes. For disciplinary knowledge, students will be able to narrate a complete, logical, coherent, and informative story. In terms of disciplinary skills, students will learn to tell a story using seven story grammars, incorporate a variety of vocabulary, and construct complex sentences with connectives. These outcomes are interconnected through disciplinary practice and pedagogical approaches. In disciplinary practice, students are expected to become effective storytellers. We have adopted a workflow inspired by the sequence of literature-based language intervention activities outlined by Gillam and Ukrainetz (2006). Our pedagogical approach combines collaborative learning and teacher-directed learning. Based on this interconnected structure, we have designed several curriculum components across the four lessons, each featuring specific learning activities.

The Learning Design Studio not only serves as a platform for course design but also provides statistical data on the course's implementation. Both teachers and SLPs can use this data to assess time allocation and content coverage throughout the course. With further integrations of learning analytics, both teachers and SLPs can track the learning progress of students, particularly those with speech and language needs.

3. CONCLUSION

This presentation demonstrates a concrete way of collaboration between teachers and SLPs in school settings, from course design to evaluating students' performance, using a well-established multi-level learning design framework. Future investigations include actual implementations in schools, evaluation of acceptability of such technology-enhanced collaboration among SLPs and teachers, and extending the application to different collaborations between educators and other professionals.

REFERENCES

Armstrong, R., Schimke, E., Mathew, A., & Scarinci, N. (2023). Interprofessional practice between speech-language pathologists and classroom teachers: A mixed-methods systematic review. *Language, Speech, and Hearing Services in Schools*, *54*(4), 1358–1376. https://doi.org/10.1044/2023_LSHSS-22-00168

Education Bureau (2023, November). *Guide on school-based speech therapy services*. Education Bureau, the Government of the Hong Kong Special Administrative Region. https://sense.edb.gov.hk/uploads/page/types-of-special-educational-needs/speech-and-language-impairment/校本言語治療服務指引_11.2023.pdf

Gillam, R., & Ukrainetz, T. (2006). Language intervention through literature-based units. In T. M. Ukrainetz (Ed.), *Literate language intervention: Scaffolding PreK–12 literacy achievement* (pp. 59–94). Eau Claire, WI: Thinking Publications.

Law, N., Li, L., Farias Herrera, L., Chan, A., & Pong, T.-C. (2017). A Pattern Language Based Learning Design Studio for an Analytics Informed Inter-Professional Design Community. *Interaction Design and Architecture(s)*, *33*, 92–112. https://doi.org/10.55612/s-5002-033-005

Law, N., & Liang, L. (2020). A Multilevel Framework and Method for Learning Analytics Integrated Learning Design. *Journal of Learning Analytics*, *7*(3), 98–117. https://doi.org/10.18608/jla.2020.73.8

Mathers, A., Botting, N., Moss, R., & Spicer-Cain, H. (2024). Collaborative working between speech and language therapists and teaching staff in mainstream UK primary schools: A scoping review. *Child Language Teaching and Therapy*, *40*(2), 120–138.

https://doi.org/10.1177/02656590241232613