The 10th International Conference on

AI, Computational Thinking and STEM Education

25 - 27 June 2026

2026

AI-CTE-STEM

Conference Proceedings

熊本大学
Kumamoto University

KYOTO UNIVERSITY
KYOTO JAPAN
FOUNDED 1897

Proceedings of the 10th International Conference on Artificial Intelligence, Computational Thinking and STEM Education

25-27 June 2026

Kumamoto

Organized by

Kumamoto University

Co-organized by

Kyoto University

National Taiwan Normal University

Publication of The Education University of Hong Kong

10 Lo Ping Road, Tai Po, New Territories, Hong Kong SAR

ISSN 2664-5661

2026
AI -CTE-STEM

Editors

Ting Chia HSU
National Taiwan Normal University, Taiwan

Hiroaki OGATA
Kyoto University, Japan

Preface

International Conference on Artificial Intelligence, Computational Thinking and STEM Education (AI-CTE-STEM 2026) is the tenth international conference, continuing from the success of the previous nineth international conferences on Computational Thinking and STEM Education. AI-CTE-STEM 2026 is organized by Kumamoto University and co-organized by Kyoto University and National Taiwan Normal University.

AI-CTE-STEM 2026 is held on 25-27 June 2026. Days 1, 2 and 3 of the conference are all held at Kumamoto Prefectural Exchange Center – "Parea Hall". The conference is the most remarkable event of the Program for worldwide sharing of ideas as well as dissemination of findings and outcomes on the implementation of artificial intelligence, computational thinking and STEM education development.

The conference this year includes keynote speeches and paper presentations.

Conference Organization:

Conference Chair

Hiroaki OGATA	Kyoto University, Japan

Program Chair

Ting-Chia HSU	National Taiwan Normal University, Taiwan

Program Co-Chair

Daner SUN	The Education University of Hong Kong
Jinbo ZHANG	Beijing Normal University, China
Juan ZUOU	Tokyo City University, Japan
Ivica BOTICKI	University of Zagreb, Croatia

International Program Committee Members *(Surnames in alphabetical order)*

Ankur BIST	Graphic Era Hill University
Shao-Chen CHANG	Yuan Ze University
Chih-Hung CHEN	National Taichung University of Education
Guang CHEN	Beijing Normal University
Jun-Ming CHEN	Soochow University
Ming-Puu CHEN	National Taiwan Normal University
WeiDong CHEN	Suzhou University of Science and Technology
Yen-Lin CHIU	National Taiwan University of Science and Technology
Chih-Ming CHU	National ILan University
Sungwan HAN	Gyeong-in National University of Education
Marc JANSEN	University of Applied Sciences Ruhr West
Hi-Lian Jeng	National Taiwan University of Science and Technology
Tai-Chien KAO	National Dong Hwa University
Chia-Pin KAO	Southern Taiwan University of Science and Technology
Zuheir N. Khlaif	An Najah National University
Mi Song KIM	Western University
Oka KURNIAWAN	Singapore University of Technology and Design

Daranee Lehtonen	University of Turku
ChangYen LIAO	National Central University
Chang-Hsin LIN	National University of Tainan
Chiu Pin LIN	National Hsinchu University of Education
Hao-Chiang Koong LIN	National University of Tainan
Jia-Jiunn LO	Chung-Hwa University
Maria José MARCELINO	University of Coimbra
Marcelo MILRAD	Linnaeus University
Kuo-Liang OU	National Hsinchu University of Education
Hüseyin ÖZÇINAR	Pamukkale University
Niwat SRISAWASDI	Khon Kaen University
Pei-Shan TSAI	National Taipei University of Technology
Meng-Ping TSUEI	National Taipei University of Education
Haipeng WAN	Capital Normal University
Sally WAN	The Chinese University of Hong Kong
Chun-Wang WEI	Far East University
Sheing-Yi WU	National Tsing-Hua University
Wan-Hsuan YEN	National Taiwan Normal University

Local Organizing Committee

Co-Chair

Rwitajit MAJUMDAR	Kumamoto University, Japan
Shin-Ichiro KUBOTA	Kumamoto University, Japan

Member

Toshihiro KITA	Kumamoto University, Japan
Yoshiko GODA	Kumamoto University, Japan
Asuka KAWAGOE	Kumamoto University, Japan
Kyoko SHIGA	Kumamoto University, Japan
Masashi TODA	Kumamoto University, Japan
Hiroshi NAKANO	Kumamoto University, Japan
Ai YAMASHITA	Kumamoto University, Japan

Conference Theme:

Foundations, Integration & Infrastructure

- AI and Computational Thinking and its Key Elements
- AI and Computational Thinking as Method
- AI and STEM and Interdisciplinary Integration
- AI and Open-Source Software and Hardware for CT and STEM Education

K–12 Classroom

- AI and Computational Thinking and Unplugged Activities in K–12
- AI and Computational Thinking and Coding Education in K–12
- AI and Computational Thinking and Subject Learning and Teaching in K–12

Higher Education & Teacher Development

- AI and Computational Thinking and Teacher Development
- AI and Computational Thinking Development in Higher Education
- AI and Computational Thinking and Evaluation

Modalities & Context

- AI and Computational Thinking and STEM/STEAM Education
- AI and Computational Thinking and Non-formal Learning
- AI and Computational Thinking and IoT
- AI and Computational Thinking and Data Science
- AI and Computational Thinking and Artificial Intelligence Education

Inclusive & Psychology

- AI and Computational Thinking and Special Education Needs
- AI and Computational Thinking and Psychological Studies

Policy & General

- AI and Computational Thinking in Educational Policy
- AI and General Submission to Computational Thinking Education

The conference received a total of 50 submissions (36 full papers, 7 short papers and 7 poster papers) from 13 countries/regions (see Table 1).

Table 1: Distribution of Paper Submissions for AI-CTE-STEM 2026

Country / Region	No. of Authors	Country / Region	No. of Authors
Taiwan	52	**Croatia**	4
Japan	22	**Indonesia**	3
Hong Kong	13	**United States**	2
South Korea	10	**Philippines**	1
Peru	8	**China**	1
Malaysia	7	**United Kingdom**	1
Singapore	5	**Total**	**129**

The International Program Committee (IPC) is formed by 35 Members and 4 Co-chairs worldwide. Each paper with author identification anonymous was reviewed by at least three IPC Members. Related sub-theme Chairs then conducted meta-reviews and made recommendation on the acceptance of papers based on IPC Members' reviews. With the comprehensive review process, 46 papers were accepted and presented at the conference, including 9 full papers, 28 short papers, and 9 poster papers (see Table 2).

Table 2: Paper Presented at AI-CTE-STEM 2026

Submission category	Submission	Accepted category	Number
Full paper	36	**Full paper**	9
		Short paper	23
		Poster	0
Short paper	7	**Short paper**	5
		Poster	2
Poster	7	**Poster**	7
Total	**50**	**Total**	**46**

The conference comprises keynote and invited speeches by internationally renowned scholars; as well as academic and poster paper presentations.

Academic and Poster Paper Presentations

There are 10 sessions of academic and poster paper presentations with 46 papers (9 full papers, 28 short papers and 9 poster papers) in the conference. Worldwide scholars present and exchange the latest research ideas and findings, which highlight the importance and pathways of computational thinking education covering K-12 education, artificial intelligence education, teacher development and STEM/STEAM education, etc.

On behalf of the Conference Organizing Committee, we would like to express our gratitude towards all speakers as well as paper presenters for their contribution to the success of AI-CTE-STEM 2026.

We sincerely hope everyone enjoys and gets inspired from AI-CTE-STEM 2026.

Prof. Ting-Chia HSU
National Taiwan Normal University, Taiwan
Program Chair of AI-CTE-STEM 2026

Prof. Daner SUN
The Education University of Hong Kong
Program Co-Chair of AI-CTE-STEM 2026

Prof. Jinbo ZHANG
Beijing Normal University, China
Program Co-Chair of AI-CTE-STEM 2026

Prof. Juan ZUOU
Tokyo City University, Japan
Program Co-Chair of AI-CTE-STEM 2026

Prof. Ivica BOTICKI
University of Zagreb, Croatia
Program Co-Chair of AI-CTE-STEM 2026

Table of Contents

Full Paper

Short Paper

Poster Paper

Hsu, T. C., & Ogata, H. (Eds.). (2026). *Proceedings of the 10th International Conference on Artificial Intelligence, Computational Thinking and STEM Education* .Kumamoto: The Asia-Pacific Society for Computers in Education.

Understanding Real-world Mathematics Problems Using AI-Generated Visual Representations: An Evaluation of a Teacher Development Course

Siu Cheung, Kong [1*], Boyuan Ren [2]

[1] Department of Mathematics and Information Technology, The Education University of Hong Kong, 10 Lo Ping Road, Tai Po, New Territories, Hong Kong

[2] Artificial Intelligence and Digital Competency Education Centre, The Education University of Hong Kong, 10 Lo Ping Road, Tai Po, New Territories, Hong Kong

[*] sckong@eduhk.hk

Abstract: *This study investigates the integration of a customised AI chatbot to support the teaching and learning of mathematics concepts through visual representations. The research examines the impact of a two-session teacher development course on the Technological Pedagogical Content Knowledge (TPACK) development of in-service teachers, enabling them to utilise an AI chatbot effectively to generate visual representations for their instruction. A total of 157 in-service mathematics teachers participated in two two-hour sessions, during which they were introduced to a customised chatbot that dynamically generates visualised teaching materials for specific real-world problems related to primary mathematics topics in Hong Kong, such as common fractions. Pre- and post-course surveys, along with course evaluations, were analysed to assess the intervention's effectiveness on teachers' TPACK components—content knowledge (CK), pedagogical content knowledge (PCK), technological content knowledge (TCK), and overall TPACK. The results indicate significant improvements in all measured domains, with teachers demonstrating increased confidence and proficiency in integrating AI-generated visual representations into their teaching practices. Despite the positive outcomes, some participants expressed concerns regarding the use of this chatbot, underscoring the need for further teacher development to bridge the gap between perceived potential benefits and practical implementation. This study highlights the role of AI in preparing teaching materials with visual representations to support students' better understanding of mathematics concepts in solving real-world problems.*

Keywords: *customised AI chatbot, mathematics, teacher development, TPACK, visual representations*

1. Introduction

External visualisation has been proven to have positive effects in science and mathematics education through visual representations, fostering deeper understanding (Schoenherr et al., 2024; Schoenherr & Schukajlow, 2024). These methods allow students to connect abstract mathematics ideas with concrete representations, bridging gaps in comprehension and improving problem-solving skills. It has been found that students occasionally face challenges in effectively constructing, using, and interpreting visualisations, raising questions about the conditions and pedagogical approaches under which visualisations are most effective (Boels et al., 2019). The integration of Artificial Intelligence (AI) in education has opened new possibilities, particularly in complex fields such as STEM education (Engelbrecht & Borba, 2024; Kong et al., 2025). This study adopted a customised AI chatbot to generate visualisations to help teachers support students' understanding of real-world mathematics problems. Beyond focusing on students' direct interactions with the customised AI chatbot, this study aimed to develop in-service teachers' Technological Pedagogical Content Knowledge (TPACK) (Koehler & Mishra, 2009) for using AI in mathematics instruction. To achieve this, two two-hour teacher development sessions were organised, during which teachers were introduced to a customised AI chatbot designed to generate visual representations of mathematics problems.

2. Literature Review: Visualisation approach in mathematics education

The integration of artificial intelligence (AI) in education has gained increasing attention in recent years, particularly in mathematics instruction (Opesemowo & Adewuyi, 2024). Technologies enable the creation of dynamic teaching and learning environments, allowing teachers to explore innovative ways to enhance students' understanding of complex mathematics concepts (Holmes et al., 2025; Kong, 2019). Among these innovations, dynamic visual approaches have proven to be an effective pedagogical strategy for enabling students' understanding of abstract mathematics concepts (Barbosa & Vale, 2021). AI tools, such as chatbots and dynamic visualisation platforms like GeoGebra, further support these strategies by automating the generation of visual representations tailored to specific learning needs (Kohen et al., 2022).

Despite the potential benefits, the effectiveness of visualisation in mathematics education remains a question. The reason for such ineffectiveness may stem from students' reluctance to use visualisation and their misinterpretation of diagrams (Boels et al., 2019), resulting in lower efficacy in learning outcomes. Some studies regard visualisation as "tools for teachers," not directly for students to use for understanding (Uesaka, 2009). Therefore, it is recommended that teachers should provide essential guidance and support during students' learning process. However, it has been found that teachers often face challenges, including a lack of technological knowledge and teacher development regarding the integration of technology-powered visualisation tools into teaching practice (Howard et al., 2021). This highlights the need for teacher development that focuses on building teachers' TPACK in using AI for mathematics teaching. This study introduces a customised AI chatbot that can stably generate visual representations for teachers to help students understand real-world mathematics problems. There is a need to provide practical examples showing how AI can assist teachers in preparing teaching materials with visualisation to enhance students' understanding.

This study aims to address this gap by investigating the impact of two two-hour teacher development sessions on in-service teachers' TPACK development, with a specific focus on using AI tools for generating visual representations to help teachers develop worksheets that guide students' understanding of real-world mathematics problems. Current chatbots without customisation cannot stably provide useful visual representations for understanding mathematics problems at a conceptual level (see Figure 1). We introduce a customised AI chatbot to teachers in this study. We illustrate how to create visualisations for designing worksheets to guide students in understanding real-world mathematics problems. The current study aims to answer the following questions: (1) What is the impact of the two two-hour teacher development sessions on in-service teachers' TPACK development? (2) How do in-service teachers perceive the effectiveness of the AI chatbot in supporting them to generate visual representations of real-world mathematics problems?

3. Methodology

3.1. Participants and instruments

This study organized two two-hour sessions for in-service mathematics teachers’ TPACK development, focusing on using a customised AI chatbot for generating visual representation to enhance students’ conceptual understanding of real-world mathematics problems. One hundred and fifty-seven in-service teachers participated in the teacher development sessions. We adopted a 5-point Likert scale survey assessing the following domains: TPACK development (12 items) and course evaluation (2 items). Among them, we received a total of

151 responses. Here are sample items in the survey of CK, PCK, TCK and TPACK respectively: CK: “I know how to support students to understand how to represent real-world problems related to fractions””; PCK: I can select and apply appropriate teaching methods—including the use of 3I (Interact, Inquiry and Invoke Thinking), AI chatbots, and peer learning—to support students to understand how to represent mathematics real-world problems”. TCK: “I understand the strengths and weaknesses of AI chatbots, and I can design appropriate prompts to support students to understand how to

represent mathematics real-world problems”; TPACK: “I can integrate the learning materials provided (such as 3I, AI chatbots and worksheets) together with proper teaching method to foster subject learning and self-regulated learning”.

3.2. An example in the course to illustrate the use of our customised AI chatbot

Figures 1 shows the images generated from a current and not customised AI chatbot with the prompt “A box contains 36 balloons. Of these, 1/3 are red and 2/9 are yellow. How many balloons are red and yellow in total? (The remaining balloons are grey.) Please display all the balloons arranged in a grid of 9 rows, with 1/3 being red, 2/9 being yellow, and the remaining balloons grey. The yellow and red balloons should be concentrated on the left side, grouped by row. Do not show any calculations, steps, or the final answer.” Current chatbots cannot stably generate the required image for teaching conceptual understanding.

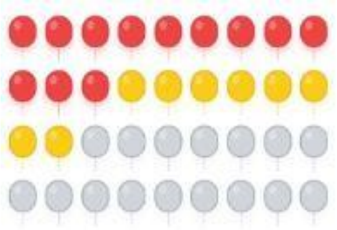


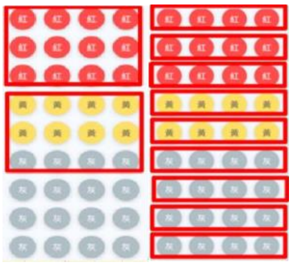
Figure 1. Current chatbot without customisation cannot stably provide a useful visual representation for conceptual understanding of mathematics problems

In the study, we designed a customised AI chatbot using RAG technique. When teachers use the prompt as shown in Step 1 of Table 1, the chatbot can stably generate the visual representation as shown on the right side of Step 1. The teachers can use the images of the two groupings to guide students to understand the equivalence of 1/3 and 3/9 using the representations on the right side of Step 2 in Table 1. Students can answer the question "A box contains 36 balloons, 1/3 of which are red, and 2/9 are yellow. How many red and yellow balloons are there in total?" by counting the balloons directly using either one of the visual representations. In this case, students cannot learn the conceptual meaning of addition of fractions. However, if teachers can design an appropriate worksheet to guide students to understand the meaning of addition of fractions using these images and this example of adding 1/3 and 2/9 and the total number of balloons of 36 to solve this problem, students can develop a genuine understanding of fraction addition.

3.3. Data analysis

The data collected from the pre- and post-course surveys were analysed to evaluate the impact of the teacher development course on in-service teachers' TPACK development. First, a reliability analysis was conducted. Cronbach's alpha was used as the measure of reliability. The overall reliability of the TPACK instrument is 0.976. The reliability analysis demonstrated good internal consistency for the survey instrument. Second, paired-samples t-tests were used to determine whether the changes in pre- and post-course scores were statistically significant. Finally, the study analysed the percentage of agreement responses to course evaluations.

Table 1. An *example of using the customised AI chatbot to create useful visual representation*

Step 1: Use a prompt to generate visual representation: "A box contains 36 balloons. Of these, 1/3 are red and 2/9 are yellow. How many balloons are red and yellow in total? (The remaining balloons are grey.) Please display all the balloons arranged in a grid of 9 rows, with 1/3 being red, 2/9 being yellow, and the remaining balloons grey. The yellow and red balloons should be concentrated on the left side, grouped by row. Do not show any calculations, steps, or the final answer."	
Step 2: Use visual representation to design a worksheet: Teachers can then use the visual representation to design a worksheet and ask students to group the balloons in two ways, one in three rows as a group and the other one row as a group. After some effort, students might be able to come up with the grouped visual representations as shown on the right side of this step.	

4. Results

The following section presents the statistical analysis results of the paired-samples t-test for in-service teachers' TPACK development before and after attending the course. The detailed analysis results are shown in Table 2 below.

Table 2: *Teachers' TPACK development regarding the use of customised AI chatbots for generating visualisations to teach conceptual understanding of real-world mathematics problems*

Items	N	Mean (SD) (pre-course)	Mean (SD) (post-course)	t	df	p
CK	151	3.40 (0.77)	3.94 (0.62)	-8.25	150	<0.001
PCK	151	3.13 (0.72)	3.88 (0.59)	-11.27	150	<0.001
TCK	151	3.00 (0.84)	3.89 (0.57)	-12.45	150	<0.001
TPACK	151	3.00 (0.80)	3.83 (0.61)	-11.46	150	<0.001
Overall	151	3.12 (0.67)	3.89 (0.56)	-12.58	150	<0.001

As shown in Table 2, the course significantly improved teachers' TPACK across all four measured domains ($p < 0.001$). The most substantial increase was in TCK, where mean scores increased from 3.00 in the pre-course to 3.89 in the post-course ($t = -12.45$, $p < 0.001$). This sharp increase highlights a marked enhancement in teachers' capacity to use technology for subject teaching. In this study, it is about the use of customised AI chatbot for teaching conceptual understanding of real-world mathematics problems. Moreover, participants showed significant gains in both foundational CK and PCK, with scores increasing from 3.40 to 3.94 ($t = -8.25$, $p<0.001$) and from 3.13 to 3.88 ($t = -11.27$, $p<0.001$), respectively. In other words, with the support of our provided teaching material on top of the customised AI chatbot, teachers developed a better understanding and pedagogy in using the generated images to support the teaching of conceptual understanding of representing real-world mathematics problems. Therefore, their overall TPACK skills also significantly increased from 3.00 to 3.83

($t = -11.46$, $p<0.001$). In summary, the consistent higher post-course scores across the four domains of TPACK underscored the effectiveness of the teacher development in enhancing teachers' TPACK components. These findings provided compelling evidence that the AI-generated visualisation together the method of designing supporting worksheets successfully supported teachers' professional growth in their teaching capability.

Additionally, we also investigated participants' evaluation of the course. The evaluation form has two items: Q1. The course improved my teaching capability on supporting students' learning of representing real-world mathematics problems; Q2. The course enhanced my confidence in supporting students' learning of representing real-world mathematics problems. The evaluation details are shown in Table 3 below.

Table 3: *Participants' evaluation after attending the two two-hour sessions*

Questions	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Mean (max 5)	SD	Total (Num)
Q1	0.6%	0.6%	21.7%	62.4%	14.6%	3.90	0.66	157
Q2	0.6%	0%	25.2%	61.9%	12.3%	3.85	0.64	155

The findings demonstrated that most teachers were satisfied with the course. In terms of the course's effectiveness, 77% of teachers agreed (62.4% agreed; 14.6% strongly agreed) that the course effectively improved their teaching capability on supporting students' learning of representing real-world mathematics problems. Similarly, the course successfully cultivated teachers' confidence in utilizing AI-generated visual representations, with 74.2% of responses. However, it was noticeable that over 30 attendees showed a neutral attitude to the questions mentioned earlier. This suggests that while the potential of the customised AI chatbot is acknowledged, further teacher development is vital to help teachers shift from theoretical understanding to practical classroom integration.

5. Discussion and conclusion

In this research, teachers serve as a vital link between technology use and student learning. The study examines how teachers can utilize a customised AI chatbot to create visual representations to support students' understanding of real-world mathematics problems which involve intricated mathematics concepts. This approach emphasizes teacher agency, allowing them to make use of customised AI chatbot to generate visual support as well as worksheet design that best support students' development of mathematics concepts through understanding of real-world mathematics problems. By focusing on the teacher's ability to curate content, this work offers a distinct alternative to the existing body of literature (Holmes et al., 2025) that primarily investigates student-facing AI applications.

Additionally, this study distinguishes itself by moving beyond the limitations of static or manually created visual representation, demonstrating the capacity of customised AI chatbots to generate dynamic teaching materials. Although this study aligned with a previous study (Kohen et al., 2022) regarding the value of dynamic visualization (e.g., GeoGebra), this research contributes to shifting from reliance on pre-built tools to empowering teachers to engage with dynamic generated visual representations actively. By introducing customised AI chatbots to teachers and guiding them to generate useful visualizations for representation of real-world mathematics problems, teachers can go back to school to use the customised AI chatbot to generate appropriate visual representations to support their teaching. This process not only can help to improve student learning outcomes but also can enhance teachers' confidence in their instructional practices.

Finally, these results supported prior research conclusions (Howard et al., 2021; Mishra & Koehler, 2009), indicating the effectiveness of teacher development in equipping teachers with the necessary skills to integrate AI into their teaching.

Despite its contributions, this study has its limitations. First, the evaluation is limited to the course's influence on teacher development. The impact of using the generated visual representations on students' learning remains underexplored. To address this, further investigations should investigate students' learning outcomes and their attitude towards the visual representations for their learning. Second, the study focuses exclusively on mathematics learning, while customised chatbots are potential pedagogical tools forteaching other subjects such Chinese language and English language, future studies should expand to include other subject teaching. This will help determine if customised AI chatbots are equally effective across different subjects in primary curriculum.

Acknowledgment

The authors would like to acknowledge the project “Using Artificial Intelligence (AI) in Primary School Education to Unleash Potential of Primary Students in Five Subjects: Chinese Language, English Language, Mathematics, Primary Science, and Primary Humanities” with the funding support from the Teaching Development Grant 2025-28 triennium, The Education University of Hong Kong.

References

Barbosa, A., & Vale, I. (2021). A visual approach for solving problems with fractions. *Education Sciences, 11*(11), 727. https://doi.org/10.3390/educsci11110727

Boels, L., Bakker, A., Van Dooren, W., & Drijvers, P. (2019). Conceptual difficulties when interpreting histograms: A review. *Educational Research Review, 28*, 100291. https://doi.org/10.1016/j.edurev.2019.100291

Engelbrecht, J., & Borba, M. C. (2024). Recent developments in using digital technology in mathematics education. *ZDM–Mathematics Education*, *56*(2), 281–292. https://doi.org/10.1007/s11858-023-01530-2

Holmes, K., Mackenzie, E., Thompson, M., Reid O’Connor, B., & Anderson, J. (2025). Increasing student interest in secondary mathematics: student perspectives. *Mathematics Education Research Journal*, 1–33. https://doi.org/10.1007/s13394-025-00553-0

Howard, S. K., Tondeur, J., Ma, J., & Yang, J. (2021). What to teach? Strategies for developing digital competency in preservice teacher training. *Computers & Education*, *165*, 104149.

Koehler, M., & Mishra, P. (2009). What is technological pedagogical content knowledge (TPACK)?. *Contemporary issues in technology and teacher education*, *9*(1), 60–70. https://doi.org/10.1016/j.compedu.2021.104149

Kohen, Z., Amram, M., Dagan, M., & Miranda, T. (2022). Self-efficacy and problem-solving skills in mathematics: The effect of instruction-based dynamic versus static visualization. *Interactive Learning Environments, 30*(4), 759–778. https://doi.org/10.1080/10494820.2019.1683588

Kong, S. C, Siu, K. Y., & Yeung, W. K. (2025). Evaluating a teacher development course for teaching STEM activities with introductory internet of things concepts and AI data model training skills using the TPACK framework: Problem-solving and digital creativity. In *Proceedings of the 17th International Conference on Computer Supported Education - Volume 1: CSEDU,* (pp. 36–47). Porto, Portugal. https://doi.org/10.5220/0013201100003932

Kong, S. C. (2019). Learning composite and prime numbers through developing an app: An example of computational thinking development through primary mathematics learning. In *Computational Thinking Education* (pp. 145–166). Singapore: Springer Singapore.

Opesemowo, O. A. G., & Adewuyi, H. O. (2024). A systematic review of artificial intelligence in mathematics education: The emergence of 4IR. *Eurasia Journal of Mathematics, Science and Technology Education*, *20*(7), em2478. https://doi.org/10.29333/ejmste/14762

Schoenherr, J., & Schukajlow, S. (2024). Characterizing external visualization in mathematics education research: A scoping review. *ZDM–Mathematics Education*, *56*(1), 73–85. https://doi.org/10.1007/s11858-023-01494-3

Schoenherr, J., Strohmaier, A. R., & Schukajlow, S. (2024). Learning with visualizations helps: A meta-analysis of visualization interventions in mathematics education. *Educational Research Review*, *45*, 100639. https://doi.org/10.1016/j.edurev.2024.100639

Uesaka, Y. (2009). How learning skills support through cognitive counseling can provide new perspectives in both cognitive research and school curriculum development: Focusing on the strategy of diagram use in problem solving. *Cognitive Studies: Bulletin of the Japanese Cognitive Science Society*, *16*(3), 313–332. https://doi.org/10.11225/jcss.16.313

Digital Transformation and Instructional Professionalism: Profiles and Predictors Among Korean Elementary Teachers Using TALIS 2024

Sang Kug NAM [1]& Hyun Sook YI [2*]
[1,2] Department of Education, Konkuk University, South Korea
*hyunsyi@konkuk.ac.kr

Abstract: *This study examines how teachers' instructional professionalism is reconfigured in the context of digital transformation. Using data from 1,879 teachers and 184 principals in Korean elementary schools from the 2024 Teaching and Learning International Survey (TALIS), Latent Profile Analysis was conducted, accounting for sampling weights and nested data structures. Random Forest and R3STEP analyses were then applied to identify factors associated with profile membership. Four distinct profiles were identified: the Digitally Integrated Professional Profile, Conventional Instructional Profile, Digitally Adaptive Generalist Profile, and Low-Engagement Profile. Key distinguishing factors included progression-based learning, adaptive instructional design, and digital self-efficacy, whereas school-level factors such as the teacher-to-administrative staff ratio showed limited influence. These findings suggest that digital professionalism is shaped more by teachers' internal pedagogical capacities than by external access alone. Consequently, the study highlights the need for tailored policy approaches that support qualitative transitions toward a more integrated form of instructional professionalism.*

Keywords: Instructional professionalism, latent profile analysis, random forest, TALIS 2024

1. Introduction

Instructional professionalism entails the professional competence required to effectively design, implement, assess, and manage classroom instruction (Lee, 2022). The COVID-19 pandemic accelerated digital transformation in education and expanded technology-enhanced teaching and learning (Oh, 2023). In response to these shifts, the South Korean government implemented the Digital-Based Education Innovation Plan, supporting the enhancement of teachers' professional competencies through digital education capacity building and the introduction of AI courseware. These shifts necessitate a reconceptualization of instructional professionalism as a construct in which digital technologies are integrated into pedagogical practice. From this perspective, the Technological Pedagogical and Content Knowledge (TPACK) framework, which emphasizes the dynamic interplay among technological, pedagogical, and content knowledge, provides a useful theoretical lens for capturing such contemporary forms of professionalism (Mishra & Koehler, 2006). However, previous studies using data from the Teaching and Learning International Survey (TALIS) 2018 have primarily focused on traditional teaching competencies, thereby limiting their ability to adequately reflect contemporary educational contexts in which digital technologies are embedded throughout the instructional process. In addition, digital education policies may influence instructional professionalism not only at the individual teacher level but also through school-level conditions, such as institutional support, leadership, and resource availability. Accordingly, this study aims to identify distinct profiles of instructional professionalism among Korean elementary school teachers by incorporating the TPACK framework and to examine the teacher- and school-level factors associated with these profiles. The specific research questions are as follows:

1. What distinct profiles of instructional professionalism emerge when digital and TPACK-related indicators are considered?
2. Which teacher-level factors are associated with membership in different instructional professionalism?

3. Which school-level factors are associated with membership in different profiles of instructional professionalism?

2. Literature review

Baek et al. (2020) utilized TALIS 2018 data to classify Korean elementary school teachers into high and low instructional professionalism groups and employed Random Forest to identify key predictors distinguishing these groups, including subject area, school climate, background characteristics, and teacher–student relationships. While digital technology use was identified as a significant predictor in their study, the scarcity of digital-related variables in TALIS 2018 restricted the ability to derive sufficient implications for teachers' digital teaching competencies. Prior studies employing Latent Profile Analysis (LPA) identified profiles of instructional professionalism among Korean secondary school teachers (Kim & Lee, 2022). However, as this research has been confined to secondary education, a gap remains in understanding how elementary school teachers' instructional professionalism is configured in relation to digital instructional practices. This gap is particularly important in the contemporary educational landscape, where digital technologies are structurally integrated into lesson design, instructional strategies, assessment, and reflection (Domínguez, 2021). This shift necessitates a reconceptualization of instructional professionalism by embedding technology use as a core component that determines the quality of instructional practice. From a TPACK perspective, this reconceptualization can be understood through the interaction among technological knowledge, pedagogical knowledge, and content knowledge, with patterns that vary qualitatively depending on the individual teacher and instructional context (Li & Bai, 2025). Aligning with this trajectory, TALIS 2024 incorporates the instructional use of digital resources and beliefs about digital technologies as domains closely tied to instructional professionalism (OECD, 2025). However, prior studies have been limited in capturing digitally integrated instructional professionalism because previous TPACK research often focused on isolated or subject-specific contexts (Baek et al., 2020; Kim & Lee, 2022; Siddiq et al., 2016). Therefore, instructional professionalism should be understood as heterogeneous profiles that combine teaching practices and digital technology use in distinct ways.

3. Data and Methods

3.1. Data

This study used TALIS 2024 data to identify instructional professionalism profiles among Korean elementary school teachers. To this end, indicator variables for the LPA were selected from the Teaching Practices and Use of Digital Resources and Tools domains in TALIS 2024. The Teaching Practices domain was measured using 18 items encompassing instructional clarity, cognitive activation, classroom management, and assessment (Cronbach's α = .80–.87; assessment α = .81), while the Use of Digital Resources and Tools domain was measured using eight items focusing on whole-class instruction, individualized instruction, and assessment (α = .75–.90). The overall reliability of the indicator variables was .90. During the analysis, participants with missing responses on the indicator variables were excluded. The final analytical sample comprised 1,879 teachers and 184 principals across 184 elementary schools.

For the predictive analysis, selected variables were categorized into teacher- and school-level predictors, including TALIS-derived indices. To ensure analytical robustness, variables already used as LPA indicators or in the construction of derived indices, variables with missing values exceeding 20%, and variables with zero variance were excluded. Remaining missing values were imputed using the mode for categorical variables and the median for continuous variables (Baek et al., 2020). Consequently, 551 predictors were retained for analysis, comprising 213 teacher-level predictors, including 7 derived variables, and 338 school-level predictors, including 17 derived variables.

3.2. Statistical Analysis

To identify distinct profiles of teachers' instructional professionalism, LPA was conducted using Mplus 7.0 (Muthén & Muthén, 1998–2012). Teacher sampling weights were incorporated to reflect the complex sampling design of the TALIS data, and standard errors and fit statistics were adjusted to account for the nested structure of teachers within schools. To determine the optimal number of latent profiles, models were estimated by incrementally increasing the number of profiles starting from a two-profile solution. Model selection was based on a comprehensive evaluation of information criteria (AIC, BIC, and SABIC), entropy, class proportions, and interpretability. The adjusted Lo–Mendell–Rubin likelihood ratio test (LMR-LRT) was used as a supplementary criterion because likelihood ratio tests may perform less well in mixture models incorporating sampling weights (Vermunt & Magidson, 2007).

Subsequently, Random Forest analysis was conducted to identify factors predicting teachers' instructional professionalism. To account for the nested structure of the TALIS data and prevent data leakage between the training and validation sets, 5-fold StratifiedGroupKFold cross-validation was applied. Model hyperparameters were optimized using GridSearchCV with Macro-F1 as the objective function, and teacher sampling weights were applied throughout the model training and validation processes. Given the class imbalance in the multi-class classification setting, predictive performance was evaluated using Macro-Precision, Macro-Recall, Macro-F1, and the Matthews correlation coefficient (MCC). All analyses were conducted in Python 3.12.2 using scikit-learn 1.5.2, numpy 1.26.4, and pandas 2.2.2.

4. Results

4.1. Teachers' instructional professionalism and its characteristics

LPA was conducted to identify profiles of instructional professionalism among Korean elementary school teachers. Model fit indices are presented in Table 1. The information criteria values consistently decreased as the number of classes increased, and entropy values exceeded .90 across all models, indicating high classification quality. Although the LMR-LRT was statistically significant only for the two-class model, previous research suggests relying on a holistic evaluation when using sampling weights. In terms of interpretability, solutions with up to four latent classes allowed for a clear differentiation of class-specific characteristics. In contrast, the five-class solution exhibited substantial conceptual overlap between profiles. Consequently, considering the balance of model fit, classification quality, and theoretical interpretability, the four-class solution was selected as the optimal model. Across all models, no latent class accounted for less than 1% of the sample. Figure 1 illustrates the distinct characteristics of the four identified latent classes.

Table 1. *Model fit indices for latent profile models of teachers' instructional professionalism*

Category		Number of latent profile classes			
		2-class	3-class	4-class	5-class
Information criteria	AIC	103468.58	100294.26	97959.96	96234.78
	BIC	103906.12	100881.34	98696.58	97120.94
	SABIC	103655.14	100544.58	98274.04	96612.62
Entropy		0.924	0.924	0.930	0.941
LMR-LRT		0.0000	0.4560	0.4471	0.2294
Proportion by group	Class 1	64.2%	30.5%	17.0%	18.7%
	Class 2	35.8%	39.4%	38.6%	29.6%
	Class 3		30.1%	21.2%	10.3%
	Class 4			23.2%	33.5%
	Class 5				7.8%

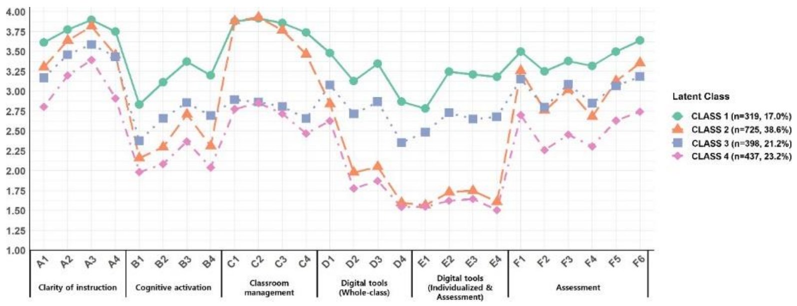


Figure 1. Characteristics of teachers' instructional professionalism.

Class 1 (17.0%), labeled the Digitally Integrated Professional Profile, showed the highest levels across both pedagogical and digital domains. Class 2 (38.6%), designated as the Conventional Instructional Profile, demonstrated high competence in instructional clarity and classroom management but showed remarkably low engagement with digital resources, suggesting a preference for traditional pedagogical methods over technological integration. Class 3 (21.2%), labeled the Digitally Adaptive Generalist Profile, displayed average levels across all indicators. Although their pedagogical scores were moderate compared to the high-performing groups, they maintained a moderate level of digital resource utilization. Finally, Class 4 (23.2%), labeled the Low-Engagement Profile, demonstrated the lowest levels across all measured domains, reflecting limited engagement in both pedagogical practices and digital tool usage.

4.2. Random forest classification results and predictors

Classification performance on the test set yielded a Macro-Precision of 0.54, a Macro-Recall of 0.51, a Macro-F1 of 0.51, and an MCC of 0.34. These metrics indicate performance above chance when equal importance was assigned to each latent class, although some overlap across classes was observed. According to the Random forest results (Table 2), key predictors included progression-based learning, self-efficacy and beliefs in digital resources, allocation of instructional time, adaptive learning, self-efficacy in classroom management, and the school-level teacher-to-administrative or management personnel ratio. The top 15 predictors were subsequently included as covariates in the R3STEP procedure.

Table 2. *Top 15 Most Important Predictors of teachers' instructional professionalism*

Rank	Item	Scale	Importance
1	I prepare students for difficulties that can occur while practicing a procedure or skill	4-point	0.013
2	Learn to use technology that is new to me [2]	4-point	0.011
3	The use of digital resources and tools helps improve students' academic performance [3]	4-point	0.011
4	Communicate with parents using digital resources and tools [2]	4-point	0.011
5	I let students review multiple examples to practice the steps involved in a procedure or skill	4-point	0.011
6	Actual teaching and learning [4]	Open-ended	0.009
7	I let students practice similar tasks until I know that every student has understood the subject matter [1]	4-point	0.009
8	I adapt my teaching methods to students' needs [5]	4-point	0.009
9	I select tasks for student practice that gradually increase in difficulty [1]	4-point	0.009
10	The use of digital resources and tools helps students collaborate on tasks efficiently [3]	4-point	0.009
11	Keeping order in the classroom, maintaining discipline [4]	Open-ended	0.008

12	Adapt the use of digital resources and tools to different teaching activities[2]	4-point	0.008
13	Total number of teachers divided by the sum of administrative and management personnel[6]	Derived ratio	0.008
14	Support student learning through the use of digital resources and tools[7]	4-point	0.008
15	I point students to different materials for learning depending on their needs[5]	4-point	0.008

Notes. Domain names: 1. Progression-based learning, 2. Self-efficacy (digital resource), 3. Beliefs in digital resources, 4. Allocation of instructional time, 5. Adaptive learning, 6. Teacher to administrative or management personnel ratio, 7. Self-efficacy (classroom management)

4.3. R3STEP with the Top 15 Predictors

Table 3 presents the results of the R3STEP analysis examining the effects of the top 15 predictors identified by the Random forest on latent profile membership, with Class 1 as the reference group. For brevity, only significant predictors are reported. Overall, high levels of progression-based learning, adaptive instructional strategies, and self-efficacy and beliefs regarding digital resources were consistently associated with a higher likelihood of belonging to Class 1 compared to the other profiles. For Class 2, lower scores on digital self-efficacy and progression-based learning significantly decreased the likelihood of membership in Class

1. However, a greater proportion of instructional time devoted to actual teaching and learning was associated with a slightly higher likelihood of belonging to Class 2 relative to the Digitally Integrated profile. This reinforces the characterization of Class 2 as a group that prioritizes traditional instructional time over digital integration. For Class 4, the odds ratios for most key predictors, including preparing students for difficulties, digital collaboration, and supporting learning through digital tools, were significantly lower compared to the reference group.

Table 3. *Effects of predictors on latent class membership*

Item	Reference Class: CLASS 1		
	CLASS 2	**CLASS 3**	**CLASS 4**
	OR	OR	OR
I prepare students for difficulties that can occur while practicing a procedure or skill[1]	-	-	0.47**
Learn to use technology that is new to me[2]	0.67*	-	0.62*
The use of digital resources and tools helps improve students' academic performance[3]	0.50***	-	-
Communicate with parents using digital resources and tools[2]	-	-	0.63**
Actual teaching and learning[4]	1.03**	-	-
I let students practice similar tasks until I know that every student has understood the subject matter[1]	-	0.60**	0.48***
I select tasks for student practice that gradually increase in difficulty[1]	0.62**	-	0.63*
The use of digital resources and tools helps students collaborate on tasks efficiently[2]	0.52**	0.53**	0.34***
Keeping order in the classroom, maintaining discipline[4]	-	0.95**	0.95**
Support student learning through the use of digital resources and tools[7]	0.61**	-	0.48***
I point students to different materials for learning depending on their needs[5]	0.68*	0.70*	0.54**

Notes. $^{*}p$<.05, $^{**}p$<.01, $^{***}p$<.001; Domain labels follow Table 2.

5. Conclusion

This study used TALIS 2024 data to examine how teachers' instructional professionalism is reconfigured in the context of digital transformation and to identify its profiles and predictors. The findings revealed four distinct profiles among Korean elementary school teachers: the Digitally Integrated Professional Profile, the Conventional Instructional Profile, the Digitally Adaptive Generalist Profile, and the Low-Engagement Profile. These classifications indicate that the use of digital technology reflects qualitatively different forms of instructional professionalism—where pedagogical expertise and digital competence do not always develop in tandem—rather than mere linear differences in adoption levels.

Profile membership was primarily predicted by progression-based learning, adaptive instructional design, and self-efficacy and beliefs related to digital resource use, whereas school-level organizational factors showed limited influence, suggesting that digital instructional professionalism may be shaped more strongly by teachers' internal capacities, such as beliefs, readiness, and technology-related competencies, than by external environments. In the Korean context, this finding is particularly relevant because recent digital education policies have broadened the roles expected of teachers. Under the Digital-Based Education Innovation Plan, teachers are increasingly expected not only to design and deliver individualized instruction using AI and digital technologies, but also to support students' social and emotional development and foster higher-order competencies through project-based learning (Ministry of Education, 2023). Accordingly, the growth of teachers' professionalism should be understood not merely as an improvement in technological skills, but as a broader transformation involving pedagogical beliefs, instructional design capacities, and student-centered support. Consequently, this highlights the need for tailored professional development programs that go beyond the mere provision of material resources to foster changes in teachers' pedagogical beliefs and support the meaningful integration of technology with subject content, thereby helping teachers in different profiles advance toward a more integrated form of professionalism.

This study has two limitations. First, the hierarchical structure of teachers nested within schools could not be fully accounted for due to convergence issues in the multilevel LPA. Second, because this study used cross-sectional TALIS 2024 data, it is limited in explaining changes in instructional professionalism over time. Future research should apply analytical approaches that better reflect the hierarchical structure of the data and explore the longitudinal development of instructional professionalism.

References

Baek, S. G., Nam, N. R., Cho, S. J., Yang, H. W. & Kim, Y. K. (2020). Exploring predictors of teachers' instructional competence using TALIS 2018 data: A comparison between elementary and middle school teachers. *The Journal of Korean Education, 47*(3), 33-63.

Galindo-Domínguez, H., & Bezanilla, M. J. (2021). Digital competence in the training of pre-service teachers: Perceptions of students in the degrees of early childhood education and primary education. *Journal of Digital Learning in Teacher Education*, *37*(4), 262–278.

Kim, T. W., & Lee, Y. J. (2022). Classification of middle school teachers' instructional practice types and exploration of predictive factors: Focusing on TALIS 2018. *Journal of Teacher Education, 39*(1), 169–202.

Lee, O. H. (2022). Differences between teachers' and students' perceptions of instructional professionalism. *Korean Journal of Educational Discourse*, *43*(1), 125–152.

Li, J., & Bai, B. (2025). Profiling TPACK in early childhood education: A multilevel latent profile analysis of teachers and preschools. *Education and Information Technologies, 30*, 26827–26855.

Ministry of Education. (2023). Digital-based education innovation plan [In Korean]. Ministry of Education.

Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record: The Voice of Scholarship in Education, 108*(6), 1017–1054.

Muthén, L. K., & Muthén, B. O. (1998–2017). *Mplus user's guide* (7th ed.). Muthén & Muthén. OECD (2025). *Teaching and Learning International Survey (TALIS) 2024 conceptual framework*.

OECD Publishing.

Oh, J. E. (2023). COVID-19 and the digital transformation of instructional methods. *Journal of Edutainment Research, 5*(3), 35–47.

Siddiq, F., Hatlevik, O. E., Olsen, R. V., Throndsen, I., & Scherer, R. (2016). Taking a future perspective by learning from the past: A systematic review of assessment instruments that aim to measure primary and secondary school students' ICT literacy. *Educational Research Review, 19*, 58–84.

Vermunt, J. K., & Magidson, J. (2007). Latent class analysis with sampling weights: A maximum-likelihood approach. *Sociological Methods & Research, 36*(1), 87–111.

Effects of Integrating Image-Based GenAI with Argumentation-based instruction, and Embodied cognitive learning on students' learning effectiveness and computational thinking skills

Ying-Chieh LIN[1], Ting-Chia HSU[2*]
[1,2] National Taiwan Normal University, Taiwan
* ckhsu@ntnu.edu.tw

Abstract: *This study investigates the effects of integrating image-based generative AI, argumentation-based instruction, and embodied cognitive learning on junior high school students' learning effectiveness and computational thinking (CT). Twenty-one students participated in a two-day AI learning program involving conceptual instruction, hands-on GenAI activities, and argumentation-supported game-based learning. Learning effectiveness and CT were assessed through pre- and posttests. The results showed significant improvements in learning effectiveness and overall CT, particularly in abstraction and decomposition. These findings indicate that combining embodied interaction with reflective argumentation can effectively support students' understanding of abstract AI concepts and the development of foundational CT skills, offering empirical support for the integration of generative AI in junior high school computer science education.*

Keywords: GenAI, argumentation, embodied cognitive, computational thinking skills

1. Introduction

With the rapid advancement of technology, artificial intelligence (AI), particularly generative AI (Gen AI), has been increasingly integrated into educational contexts, offering new possibilities for instructional design and learning processes (Su & Yang, 2023; Holmes & Tuomi, 2022). Prior studies have demonstrated the potential of Gen AI to enhance learning outcomes and reduce learning anxiety across disciplines such as language and mathematics (Hsu et al., 2024; Wardat et al., 2023). However, empirical research on the application of Gen AI in computer science education remains limited. To address this gap, the present study integrates Gen AI into computer science instruction to support student learning.

Computer science education often involves highly abstract concepts that pose challenges for middle school students. To facilitate students' understanding, this study adopts argumentation-based instruction, which provides structured guidance for evaluating multiple perspectives and supports the development of independent thinking (Zheng et al., 2023). The instructional design is grounded in the Toulmin Argumentation Model, a systematic framework that helps students organize and reason about abstract concepts (Rismanto et al., 2021). Nevertheless, argumentation alone may impose excessive cognitive demands on learners. Although middle school students have entered Piaget's formal operational stage, they still require concrete experiences to effectively internalize abstract concepts such as programming and AI (Piaget, 1972). Therefore, this study incorporates embodied cognition to complement argumentation-based instruction. Embodied cognition emphasizes the role of bodily experience and environmental interaction in cognitive processes, making it particularly suitable for supporting middle school students in transforming abstract concepts into concrete and actionable knowledge (Kersting et al., 2021).

Focusing on knowledge-level learning outcomes (Ji et al., 2022) and computational thinking (CT)—a core interdisciplinary competence in computer science education (Kong, 2019)—this study designs guided and interactive Gen

AI–supported instructional activities. The study examines the effects of this instructional approach on students' learning outcomes and CT skills, aiming to provide empirical evidence and instructional design implications for the integration of Gen AI in computer science education.

2. Literature review

2.1. Argumentation-based instruction： Toulmin Model

The Toulmin Argument Model, comprising claim, data, warrant, backing, qualifier, and rebuttal, is regarded as the most comprehensive framework for analyzing argument structure (Rismanto et al., 2021) and has been widely applied in cross-disciplinary instruction. In education, structured argumentation fosters deep understanding, multi-perspective thinking, and improved reasoning quality (Nussbaum, 2021). Recent advances in AI have spurred interest in integrating AI with argumentation, with studies showing AI can act as a virtual learning partner, provide feedback in inquiry or writing, and, when combined with gamified learning, boost engagement and perspective diversity (Lin & Hung, 2025). However, despite its analytical strengths, the Toulmin model lacks standardized criteria for evaluating argument quality, posing challenges for novice learners (Nussbaum, 2021).

2.2. Embodied cognition with AI education

Embodied cognition theory emphasizes that cognition is not solely the result of brain activity, but rather emerges from the interaction among the brain, the body, and the environment, and is influenced by bodily movements, postures, and interactions with the surrounding context (Castro-Alonso et al., 2024). In a broad sense, embodiment focuses on the bodily experiences accumulated through individuals' interactions with the material and sociocultural world, whereas embodied cognition further explains how the body actively participates in thinking and cognitive processes (Popova & Rączaszek-Leonardi, 2020; Kersting et al., 2021).

Related studies indicate that through concrete manipulation and bodily engagement, learners can use sensory and motor experiences as a foundation for understanding abstract concepts. This perspective aligns with Piaget's (1972) theory of cognitive development and Dewey's (1938) emphasis on learning as an interaction between action and reflection. Barsalou's (1999) perceptual symbol systems theory also suggests that when humans comprehend abstract concepts, they reactivate perceptual representations associated with concrete experiences. Recent research has shown that embodied cognition learning can effectively enhance students' understanding of abstract concepts in programming and computer science (Kwon et al., 2022). However, most existing studies have focused on adults or university students, with relatively limited exploration of the effects of different levels of embodiment on adolescents' AI learning outcomes (Johnson-Glenberg, 2014). Therefore, this study aims to compare the effects of AI course designs with different levels of embodiment on junior high school students' learning achievement, CT, and metacognitive awareness.

3. Methodology

3.1. Instructional Materials

This study employed the newly developed AI 2 Robot City board game, created by Professor Ting-Chia Hsu's team from National Taiwan Normal University. The educational board game integrates image recognition and Internet of Things (IoT) technologies to help students effectively enhance their CT abilities through game-based learning (GBL).

On the software side, the MIT App Inventor platform developed by the Massachusetts Institute of Technology was used to design the mobile application. This was combined with the Personal Image Classifier (PIC) web platform for implementing image recognition functionalities. After training a custom model on the PIC platform, users can import the trained model into the App Inventor environment. The model can then be used to control the robot car based on image input. The board game is played in a mission-based format, with two students per group and two groups competing against each other. Players earn points by collecting building materials shown on the mission cards, with each card offering different point values. This encourages students to collaborate and strategize on how to maximize their score within a limited time. The group with the highest score at the end of the game is declared the winner, as shown in Figure 1.

Figure 1 AI 2 Robot City board game

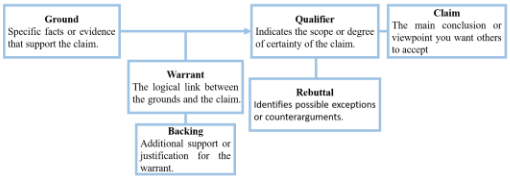


Figure 2 Toulmin Argument Model

3.2. Teaching strategy

To investigate the effects of embodied cognition and argumentation-based instruction on students' learning effectiveness, this study designed an integrated instructional approach centered on embodied cognition. The argumentation framework, based on the Toulmin Argument Model, was implemented through a circular worksheet to support students in constructing claims and articulating reasoning through grounds, backing, and rebuttals, thereby externalizing their thinking processes, as shown in Figure 2.

Embodied cognition was operationalized through three activities: a "deep learning neural network simulation," where students physically arranged images to represent AI classification processes; a "generative AI card game," where they modified prompt sequences to generate different outputs; and a "board game activity," where students engaged in collaborative decision-making through the AI2 Robot City game, integrating physical interaction with cognitive processes.

Argumentation-based strategies further supported all activities. The first two activities used worksheets to guide written reflection, while the board game emphasized peer discussion, argumentation, and consensus-building, enhancing students' reasoning and critical thinking skills, as shown in Table 1.

Table 1 *Instructional Design Framework*

	Instructional Design	Integration with Argumentation
Activity 1: Deep Learning Neural Network Simulation	By physically moving and arranging images, students simulate how AI performs image classification.	Post-activity argumentation worksheet to promote concrete thinking and reduce cognitive load.
Activity 2: Gen AI Game Cards	Students rearrange paper cards to change the order of the prompt, generate images, and draw the results.	Post-activity argumentation worksheet to promote reflection and awareness in claim construction and justification.
Activity 3: Board Game Playing	The AI2 Robot City board game enables students to integrate embodiment into tabletop gameplay activities.	Oral discussion with embodied actions to promote peer argumentation, rebuttal, and consensus building.

3.3. Teaching strategy

This study examined the effects of integrating image-based generative AI, argumentation-based instruction, and embodied cognitive learning, along with game-based learning (GBL), on junior high school students' learning effectiveness and CT. Twenty-one students (aged 12–15) from northern Taiwan participated in a two-day AI learning camp designed to support understanding of abstract AI concepts while fostering cognitive and embodied learning processes. Although the sample size was small due to voluntary recruitment, the intensive hands-on design required close guidance and small-group facilitation. Therefore, this study serves as an exploratory pilot providing initial empirical evidence. Overall, argumentation-based instruction functioned as a framework guiding students from task execution to deeper reflection, reasoning, and collaboration.

On Day 1, students completed a pre-test, studied foundational AI and deep learning concepts, and used the Toulmin argumentation model to visualize reasoning. They generated images via Leonardo AI, translated prompts with Google Translate, and manually redrew outputs to address AI ethics. Students then used the PIC platform to create datasets, train image recognition models, and export them.

On Day 2, students engaged in block-based programming with MIT App Inventor to integrate their models with robotic systems, refined their models following a mini-lecture on optimization, and completed Argumentation Worksheet 2 for reflective reasoning. The program concluded with a custom board game in which teams applied their models to complete mission-based tasks, reinforcing CT, problem-solving, and collaboration. The experimental procedure is shown in Figure 3.

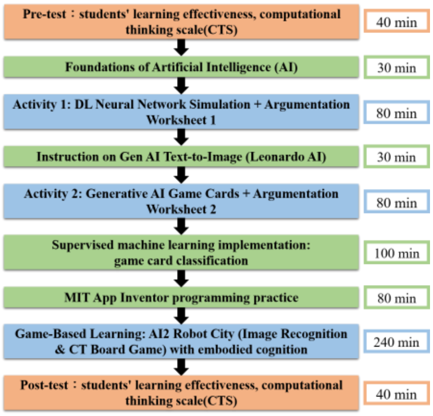


Figure 3 Experimental Procedure

3.4. Experimental tool

To examine whether the incorporation of embodied cognition influences students' learning effectiveness, this study employed three quantitative instruments：

First, learning effectiveness was measured using a 20-item test (17 multiple-choice, 3 matching; score range 0–100). Computational Thinking Scale (CTS) was assessed using the five-dimensional scale developed by Tsai et al. (2020), which includes abstraction, decomposition, algorithmic thinking, evaluation, and generalization. The Cronbach's alpha values for each dimension were .81, .74, .77, .83, and .75, respectively. For example, an item measuring abstraction is: "I usually try to identify the key points of a problem."

4. Conclusions

4.1. Student Learning Effectiveness

A comparison of students' performance showed significant improvement from the pretest to the posttest. This indicates that combining argumentation-based instruction with AI-based image generation in game-based learning can significantly enhance students' learning effectiveness. Table 2 presents the summary results of the paired-sample t-test for students' pre- and posttest progress scores.

Table 2 *Summary table of paired t-test for students' learning effectiveness*

	N	Mean	*SD*	*t*	*df*
Pretest	21	79.76	9.43	7.77**	20
Posttest	21	89.24	9.17		

***p < .01*

4.1. Student Learning Effectiveness

The overall CTS results showed significant improvement from the pretest to the posttest, indicating that combining argumentation-based instruction with AI-based image generation in game-based learning can significantly enhance students' CT. Specifically, significant improvements were observed in the Abstraction, Decomposition, Algorithmic Thinking, and Generalization dimensions based on the paired-sample t-test results. Table 3 presents the summary results of the paired-sample t-test for students' pre- and posttest progress scores.

Table 3 *Summary table of paired t-test for Computational Thinking Scale(CTS) dimension*

Dimension	*N*	Pretest Mean (SD)	Posttest Mean (SD)	*t*	*df*
Abstraction	21	3.63 (0.57)	4.04 (0.70)	4.05***	20
Decomposition	21	3.52 (0.61)	3.86 (0.56)	2.60*	
Algorithmic Thinking	21	3.49 (0.60)	4.08 (0.66)	3.32**	
Evaluation	21	3.69 (0.68)	3.98 (0.66)	1.75	
Generalization	21	3.43 (0.73)	3.99 (0.68)	3.28**	
Overall	21	3.55 (0.50)	4.00 (0.54)	3.51***	

**p < .05, **p < .01, ***p < .001*

5. Discussion and Conclusion

The results of this study indicate that integrating image-based generative AI, embodied cognitive learning, argumentation-based instruction, and game-based learning can effectively enhance junior high school students' learning effectiveness and overall computational thinking. Significant improvements were found in Abstraction, Decomposition, Algorithmic Thinking, and Generalization, suggesting that the instructional design not only supported students' understanding of abstract AI concepts but also promoted their problem analysis, procedural planning, and knowledge transfer abilities.

These findings may be related to the design of the learning activities. Through image arrangement, card manipulation, model training, App Inventor programming, and board game tasks, students transformed abstract concepts into concrete operational experiences, which is consistent with embodied cognition perspectives emphasizing the role of bodily experience and interaction in learning abstract concepts (Kersting et al., 2021). At the same time, argumentation worksheets and peer discussions guided students to explain their reasoning, reflect on strategies, and apply what they had learned, aligning with prior research suggesting that structured argumentation can support reasoning, reflection, and deeper understanding (Nussbaum, 2021). Therefore, the combination of concrete manipulation, reflective argumentation, and task-based application may have contributed to the development of multiple CT dimensions. However, the Evaluation dimension did not show significant improvement. This may be because evaluation skills involve higher-order abilities, such as judging the quality of solutions, comparing different strategies, debugging, and reflective decision-making, which usually require longer practice and more explicit instructional scaffolding (Tsarava et al., 2022). Therefore, future research may extend the instructional

duration and incorporate more testing, debugging, comparison, and reflection activities to further support students' evaluation skills.

Overall, this study suggests that GenAI-based instruction integrating embodied cognition, argumentation-based instruction, and game-based learning can effectively support students' development in most CT dimensions and provide implications for AI curriculum design in junior high school education.

References

Castro-Alonso, J. C., Ayres, P., Zhang, S., de Koning, B. B., & Paas, F. (2024). Research avenues supporting embodied cognition in learning and instruction. *Educational Psychology Review*, *36*(1), 10. https://doi.org/10.1007/s10648-024-09847-4

Holmes, W., & Tuomi, I. (2022). State of the art and practice in AI in education. *European journal of education*, *57*(4), 542-570. https://doi.org/10.1111/ejed.12533

Hsu, T. C., Chang, C., & Jen, T. H. (2024). Artificial intelligence image recognition using self-regulation learning strategies: effects on vocabulary acquisition, learning anxiety, and learning behaviours of English language learners. *Interactive Learning Environments*, 32(6), 3060-3078. https://doi.org/10.1080/10494820.2023.2165508

Ji, M., Luo, Z., Feng, D., Xiang, Y., & Xu, J. (2022). Short- and long-term influences of flipped classroom teaching in physiology course on medical students' learning effectiveness. *Frontiers in Public Health*, 10. https://doi.org/10.3389/fpubh.2022.835810

Johnson-Glenberg, M. C., Birchfield, D. A., Tolentino, L., & Koziupa, T. (2014). Collaborative embodied learning in mixed reality motion-capture environments: Two science studies. *Journal of educational psychology*, 106(1), 86. https://doi.org/10.1037/a0034008

Kersting, M., Haglund, J., & Steier, R. (2021). A growing body of knowledge: On four different senses of embodiment in science education. *Science & Education*, *30*(5), 1183-1210. https://doi.org/10.1007/s11191-021-00232-z

Kong, S. C. (2019). Components and methods of evaluating computational thinking for fostering creative problem-solvers in senior primary school education. In Computational thinking education (pp. 119-141). *Singapore: Springer Singapore*. https://doi.org/10.1007/978-981-13-6528-7_8

Kwon, K., Jeon, M., Zhou, C., Kim, K., & Brush, T. A. (2022). Embodied learning for computational thinking in early primary education. *Journal of Research on Technology in Education*, *56*(4), 410– 430. https://doi.org/10.1080/15391523.2022.2158146

Popova, Y. B., & Rączaszek-Leonardi, J. (2020). Enactivism and ecological psychology: The role of bodily experience in agency. *Frontiers in Psychology*, 11. https://doi.org/10.3389/fpsyg.2020.539841

Rismanto, R., Asri, A. N., Andoko, B. S., Hirashima, T., & Leonanta, A. A. (2021). A Preliminary Study: Toulmin Arguments in English Reading Comprehension for English as Foreign Language Students. In 2021 International Conference on Electrical and Information Technology (IEIT) (pp. 6-10). *IEEE*. https://doi 10.1109/IEIT53149.2021.9587398

Su, J., & Yang, W. (2023). Unlocking the power of ChatGPT: A framework for applying Gen AI in education. ECNU Review of Education, 6(3), 355-366. Su, J., & Yang, W. (2023). Unlocking the power of ChatGPT: A framework for applying Gen AI in education. *ECNU Review of Education*, *6*(3), 355-366. https://doi.org/10.1177/2096531123116842

Tsarava, K., Moeller, K., Román-González, M., Golle, J., Leifheit, L., Butz, M. V., & Ninaus, M. (2022). A cognitive definition of computational thinking in primary education. *Computers & Education*, *179*, 104425. https://doi.org/10.1016/j.compedu.2021.104425

Tsai, M.-J., Liang, J.-C., & Hsu, C.-Y. (2020). The Computational Thinking Scale for Computer Literacy Education. *Journal of Educational Computing Research*, *59*(4), 579-602. https://doi.org/10.1177/0735633120972356

Nussbaum, E. M. (2021). Critical integrative argumentation: Toward complexity in students' thinking. *Educational Psychologist*, 56(1), 1-17. https://doi.org/10.1080/00461520.2020.1845173

Lin, Y. R., & Hung, C. Y. (2025). The Synergistic Effects in an AI-Supported Online Scientific Argumentation Learning Environment. *Computers & Education*, 105251. https://doi.org/10.1016/j.compedu.2025.105251

Wardat, Y., Tashtoush, M. A., AlAli, R., & Jarrah, A. M. (2023). ChatGPT: A revolutionary tool for teaching and learning mathematics. Eurasia Journal of Mathematics, *Science and Technology Education*, *19*(7), em2286. https://doi.org/10.29333/ejmste/13272

Zheng, X. L., Huang, J., Xia, X. H., Hwang, G. J., Tu, Y. F., Huang, Y. P., & Wang, F. (2023). Effects of online whiteboard-based collaborative argumentation scaffolds on group-level cognitive regulations, written argument skills and regulation patterns. *Computers & education*, 207, 104920. https://doi.org/10.1016/j.compedu.2023.104920

Hsu, T. C., & Ogata, H. (Eds.). (2026). *Proceedings of the 10th International Conference on Artificial Intelligence, Computational Thinking and STEM Education* .Kumamoto: The Asia-Pacific Society for Computers in Education.

Emotion Categories and Learning Outcomes in OMO Learning Analytics Using Facial-Expression Recognition

Yi-Wen Liao[1], Yen-Cheng Tseng[2], Chuan-Sheng Lin[3]* & Ya-Han Liang[4]
[1,4] Department of Intelligent Commerce, National Kaohsiung University of Science and Technology, Taiwan (R.O.C.)
[2] Affiliated Senior High School of National Kaohsiung Normal University, Taiwan (R.O.C.) cDepartment of [3] Computer Science and Information Engineering, National Pingtung University, Taiwan (R.O.C.)
*chuanshenglin@nptu.edu.tw

Abstract: *Online-Merge-Offline (OMO) learning environments generate rich behavioral signals, yet instructors often lack timely and actionable indicators of learners' affective states during instruction. This study presents an OMO learning analytics workflow that integrates a web-based AI learning platform with facial-expression-based emotion recognition to examine how emotion-related signals relate to learning outcomes. The workflow captures learners' facial expressions during learning activities, classifies them into seven emotion categories, and aggregates emotion indicators for subsequent analysis together with assessment results collected in the course. Based on data from 30 participants, we analyze two complementary emotion indicators: (1) emotion change frequency and (2) emotion category distribution. The results show that emotion change frequency is not significantly correlated with learning outcomes, whereas emotion category distribution is meaningfully associated with learners' performance. These findings suggest that, in OMO learning analytics, emotion categories may provide more informative cues than simple frequency-based measures for understanding learning outcomes. The proposed workflow offers a practical foundation for emotion-aware learning analytics in OMO settings and supports future design of instructional interventions and dashboards that reflect affective information aligned with learning performance.*

Keywords: OMO learning, learning analytics, facial expression, emotion recognition, learning outcomes, AI learning platform

1. Introduction

Online learning has become a major trend in education, reshaping instructional design through internet-based delivery, mobile access, and data-driven support (Qazi et al., 2024; Gros & García-Peñalvo, 2023). Compared with traditional classroom instruction, e-learning enables flexible access to learning resources and assessments, which helps learners study beyond time and place constraints (Alenezi et al., 2023). However, in practice, many courses still rely heavily on post-hoc evaluations (e.g., end-of-unit tests or surveys). Such approaches often provide delayed feedback and make it difficult for instructors to recognize learning problems early, particularly in interactive or activity-rich learning settings (Mohamed, 2024).

In recent years, Online-Merge-Offline (OMO) learning has emerged as a common instructional arrangement, where online learning activities are integrated with face-to-face sessions. OMO settings can generate rich learning traces, yet instructors may still struggle to interpret learners' learning states in a timely and actionable manner. In addition to behavioral logs and assessment scores, learners' emotions are widely acknowledged as a key factor associated with cognition, motivation, and academic performance (Pekrun et al., 2002; Pekrun & Stephens, 2010). In other words, affective signals can complement conventional learning analytics by providing another perspective on how learners experience instructional activities. Nevertheless, emotion-related information is often difficult to obtain consistently in real teaching contexts, and simple frequency-based summaries may not adequately reflect meaningful affective patterns related to learning outcomes.

To address this gap, this study proposes an OMO learning analytics workflow that integrates a web-based AI learning platform with facial-expression-based emotion recognition, and then links emotion-related indicators to learning outcomes. In our workflow, learners' facial videos are captured during learning activities, facial regions are detected from video frames, and facial expressions are classified into seven emotion categories. The recognized emotions are then aggregated into two interpretable indicators for subsequent analysis: emotion change frequency and emotion category distribution. These indicators are analyzed together with learners' performance (learning outcomes) to examine how different emotion-related signals relate to learning effectiveness in an OMO context.

Based on data collected from 30 participants, our analysis yields two key findings. First, emotion change frequency does not show a significant correlation with learning outcomes. Second, emotion category distribution is meaningfully associated with learners' performance. These results suggest that, for emotion-aware learning analytics in OMO settings, category-level emotion information may provide more informative cues than frequency-based measures when examining associations with learning outcomes. Figure 1 illustrates the overall workflow and the locations where emotion indicators and learning outcomes are integrated for analysis.

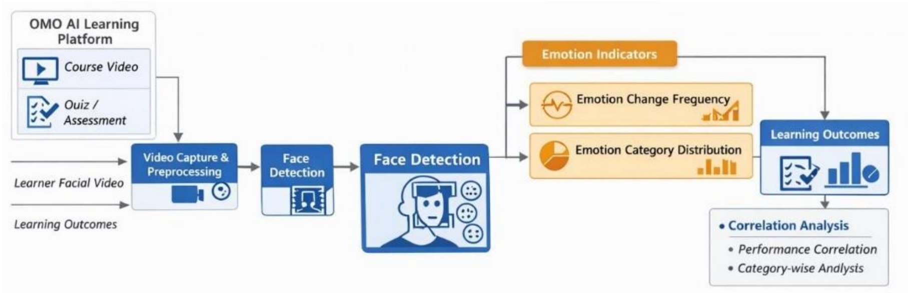


Figure 1. OMO learning analytics workflow

2. Related Work

E-learning and learning analytics. E-learning continues to expand with evolving instructional designs and digital affordances, while also facing adoption barriers and practical constraints in real teaching contexts (Qazi et al., 2024; Gros & García-Peñalvo, 2023; Sofi-Karim et al., 2023). Learning analytics leverages learning-process and behavioral data to understand and predict performance (Qiu et al., 2022), but many approaches still rely on post-hoc outcomes and may provide limited timely insight for instructors in interactive settings (Mohamed, 2024).

Academic emotions. Research indicates that learners' emotions are closely related to motivation, self-regulation, and achievement (Pekrun, 2009; Pekrun & Stephens, 2010). In online learning, academic emotions can mediate the relationship between interaction/engagement and learning outcomes, highlighting emotions as a meaningful signal beyond behavioral logs alone (Wang et al., 2022).

Facial-expression signals. Advances in computer vision and deep learning have made face-related analysis feasible for practical pipelines, supported by widely used tools and foundational methods (Bradski & Gary, 2000; LeCun et al., 1998). Representative deep face models further strengthened feature representation (Schroff et al., 2015). Building on this line, we integrate facial-expression-based emotion recognition into an OMO learning analytics workflow and examine how derived emotion indicators relate to learning outcomes.

3. Method

Study setting and data sources. This study was conducted in an Online-Merge-Offline (OMO) learning context supported by a web-based AI learning platform. The platform provided learning materials and quiz/assessment activities.

Two types of data were collected: (1) learners' facial videos during learning and (2) learning outcomes from assessment results. Data from 30 participants were included.

The workflow treats learning outcomes as the target performance measure and derives emotion indicators from facial-expression signals captured during learning; the overall data flow is summarized in Figure 1.

Emotion recognition pipeline. As shown in Figure 1, the workflow transforms raw facial videos into emotion indicators. Videos are processed into frames, faces are detected and extracted, and facial expressions are classified into seven emotion categories to form an emotion label sequence for each learner.

Figure 2 shows an example output of face detection and emotion classification, illustrating per-frame predictions before aggregation into emotion indicators.

Figure 2. Example emotion recognition output.

Emotion indicators. After obtaining the emotion label sequence for each learner, we derived two interpretable indicators: (1) emotion change frequency, measuring how often the predicted label changes over time, and (2) emotion category distribution, summarizing the overall composition of the seven predicted categories (e.g., counts or proportions). These indicators are computed per learner/session and analyzed with learning outcomes (Figure 1).

1. Emotion change frequency measures how frequently the predicted emotion label changes over time within a learning session, reflecting the temporal variability of emotion signals.
2. Emotion category distribution summarizes the overall composition of the seven predicted emotion categories during learning (e.g., counts or proportions), reflecting which categories are more prevalent during a learning session.

These two indicators are complementary: change frequency captures the dynamics of the emotion sequence, while category distribution captures the overall emotional composition. Both indicators are computed per learner/session and are later analyzed together with learning outcomes, as shown in Figure 1.

Analysis. Learning outcomes from the platform's assessment records were used as the performance measure. We conducted correlation analysis between the emotion indicators and learning outcomes (Table 1) and then performed a category-wise analysis based on emotion category distribution to examine outcome differences across emotion-category patterns (Figure 3).

Table 1. *Correlation analysis results.*

correlation	A_TEST	B_TEST	Score	Frequency	times	minutes
A_TEST	1	0.680**	0.911**	0.031	0.025	0.044
B_TEST	0.680**	1	0.922**	-0.041	-0.054	-0.040
Score	0.911**	0.922**	1	-0.007	-0.017	0.000
Frequency	0.031	-0.041	-0.007	1	0.978**	0.291
times	0.025	-0.054	-0.017	0.978**	1	0.475**
minutes	0.044	-0.040	0.000	0.291	0.475**	1

** The correlation is significant at a significance level of 0.01.

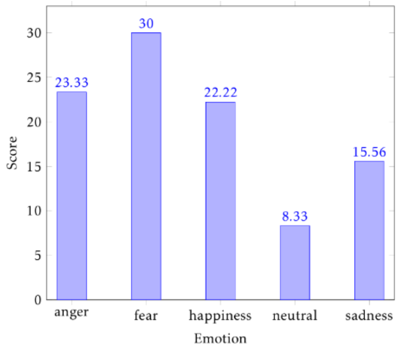


Figure 3. Category-wise results based on emotion category distribution.

4. Results and Discussion

Indicator-level results. We examined associations between the two emotion indicators (emotion change frequency and emotion category distribution) and learning outcomes using correlation analysis; Table 1 summarizes the results.

The correlation analysis indicates that emotion change frequency does not show a significant association with learning outcomes (Table 1). This finding suggests that a frequency-based summary of emotion dynamics alone may be insufficient for interpreting learning effectiveness in the OMO setting examined in this study. A high frequency of label changes may reflect short-term fluctuations during learning, but such variability does not necessarily correspond to better (or worse) assessment performance. Therefore, using emotion change frequency as a single affective indicator may provide limited actionable insight for instructors in OMO contexts.

In contrast, results related to emotion category distribution point to category-level affective information as a more meaningful signal for understanding performance. This motivates further examination of how different emotion categories relate to learning outcomes, as reported in the following subsection.

Category-wise results. We conducted a category-wise analysis based on learners' emotion category distribution. Figure 3 summarizes the results and compares learning outcomes across emotion-category patterns.

The category-wise results show that learning outcomes differ across emotion categories, indicating that emotion categories observed during learning are meaningfully associated with performance in this OMO context (Figure 3). Compared with change frequency, emotion category distribution preserves information about which affective states are prevalent during learning, and this appears to better differentiate learning outcomes.

Taken together with the correlation results in Table 1, these findings support the view that emotion category distribution is a more informative indicator than emotion change frequency when linking facial-expression-based emotion signals to learning outcomes in the proposed workflow.

From a practical perspective, this combined evidence suggests that emotion-aware learning analytics for OMO instruction may benefit from presenting category-level summaries (e.g., distribution profiles) rather than relying solely on dynamic-frequency metrics. Such summaries can be more interpretable to instructors and more directly aligned with learning outcomes, which supports reflective teaching and the design of targeted instructional support in AI-enabled learning environments.

Implications and limitations. The proposed workflow (Figure 1) provides a feasible pathway for incorporating facial-expression-based affective signals into learning analytics in technology-supported instruction. However, the evidence is based on a specific OMO learning setting with 30 participants, and per-frame emotion recognition may introduce uncertainty that influences derived indicators; therefore, generalizability to other contexts should be interpreted with caution.

5. Conclusion

This paper presented an OMO learning analytics workflow that integrates a web-based AI learning platform with facial-expression-based emotion recognition and links emotion-related indicators to learning outcomes. Using data from 30 participants, we examined two indicators derived from seven-category emotion recognition: emotion change frequency and emotion category distribution. The results show that emotion change frequency is not significantly associated with learning outcomes, whereas emotion category distribution is meaningfully associated with learners' performance (see Table 1 and Figure 3). These findings suggest that, in OMO learning analytics, category-level emotion information may provide more informative cues than frequency-based measures for understanding learning outcomes. Overall, the proposed workflow offers a practical foundation for emotion-aware learning analytics in OMO settings and supports future refinement of indicators and instructional support based on affective information.

Acknowledgements

This work was supported by the National Science and Technology Council (NSTC), Taiwan, under Grant NSTC 113-2410-H-992-031-MY2

References

Alenezi, M., Wardat, S., & Akour, M. (2023). The need of integrating digital education in higher education: Challenges and opportunities. Sustainability, 15(6), 4782.

Bradski, G., & Gary. (2000). The OpenCV Library. Dr. Dobb's Journal of Software Tools.

Gros, B., & García-Peñalvo, F. J. (2023). Future trends in the design strategies and technological affordances of e-learning. In Learning, design, and technology: An international compendium of theory, research, practice, and policy (pp. 345–367). Springer.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278–2324.

Mohamed, A. M. (2024). Exploring the potential of an AI-based chatbot (ChatGPT) in enhancing English as a foreign language (EFL) teaching: Perceptions of EFL faculty members. Education and Information Technologies, 29(3), 3195–3217.

Pekrun, R. (2009). Emotions at school. In Handbook of motivation at school (pp. 589–618). Routledge.

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. Educational Psychologist, 37(2), 91–105.

Pekrun, R., & Stephens, E. J. (2010). Achievement emotions: A control-value approach. Social and Personality Psychology Compass, 4(4), 238–255.

Qazi, M. A., Sharif, M. A., & Akhlaq, A. (2024). Barriers and facilitators to adoption of e-learning in higher education institutions of Pakistan during COVID-19: Perspectives from an emerging economy. Journal of Science and Technology Policy Management, 15(1), 31–52.

Qiu, F., Zhang, G., Sheng, X., Jiang, L., Zhu, L., Xiang, Q., Jiang, B., & Chen, P. K. (2022). Predicting students' performance in e-learning using learning process and behaviour data. Scientific Reports, 12(1), 453.

Schroff, F., Kalenichenko, D., & Philbin, J. (2015). FaceNet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 815–823). IEEE.

Sofi-Karim, M., Bali, A. O., & Rached, K. (2023). Online education via media platforms and applications as an innovative teaching method. Education and Information Technologies, 28(1), 507–523.

Wang, Y., Cao, Y., Gong, S., Wang, Z., Li, N., & Ai, L. (2022). Interaction and learning engagement in online learning: The mediating roles of online learning self-efficacy and academic emotions.

Learning and Individual Differences, 94, 102128.

Hsu, T. C., & Ogata, H. (Eds.). (2026). *Proceedings of the 10th International Conference on Artificial Intelligence, Computational Thinking and STEM Education* .Kumamoto: The Asia-Pacific Society for Computers in Education.

Preschoolers' Computational Thinking Learning: Teachers' Awareness of Stop-Motion Animation Learning Experiences in STEM Education

Yi-Hui LIN[1*], Ching-Yeh WANG[2]
[1] Center for Teacher Education and Career Development, University of Taipei, Taiwan
[2] Department of Education, National Taipei University of Education, Taiwan
*yihuilin@utaipei.edu.tw

Abstract: *Stop-motion animation has increasingly been incorporated into early childhood STEM education as a creative, design-oriented learning activity. However, limited research has examined how preschool teachers recognize the computational thinking (CT) learning experiences embedded within such activities. This exploratory qualitative study investigates preschool teachers' awareness of CT as children engaged in stop-motion animation projects within a STEM curriculum. Two animation projects, an animation centered on children's awareness of campus cultural architecture and a soccer-themed story, were implemented within the same instructional context. Data were collected through semi-structured interviews with three preschool teachers and classroom-based documentation of the animation-making processes. Thematic analysis revealed that teachers identified multiple CT dimensions as children planned stories, decomposed experiences into structured representations, coordinated frame-by-frame sequencing, refined animations through playback, and engaged in iterative revision. Teachers particularly noted how the temporal and incremental nature of stop-motion production made children's decision-making processes visible, including decomposition, sequencing, algorithmic reasoning, and debugging. The findings suggest that stop-motion animation functions as a pedagogical medium through which computational thinking becomes observable in early childhood STEM classrooms. By externalizing children's design decisions and iterative adjustments, animation-making supported teachers in recognizing CT as an experiential, process-oriented form of thinking embedded within everyday STEM learning practices.*

Keywords: Preschoolers; computational thinking; teachers' awareness; stop-motion animation; STEM education

1. Introduction

Computational thinking (CT) has become a central focus in K–12 education, particularly in discussions of how learners engage in problem solving, sequencing, and systematic reasoning across subject areas. In early childhood education, CT is increasingly understood as a way of thinking embedded within developmentally appropriate classroom activities rather than confined to formal programming instruction. Accordingly, researchers and educators have explored pedagogical approaches that allow young children to experience CT in concrete and meaningful ways. Within STEM education, creative and representational activities have been identified as promising contexts for supporting children's engagement with CT. Stop-motion animation, in particular, foregrounds sequencing, procedural planning, and revision through frame-by-frame construction. As children design animations, they make iterative decisions about movement, timing, representation, and narrative structure—processes aligned with core CT elements. Although prior research has examined children's engagement and learning outcomes in animation-based activities, less attention has been paid to how teachers recognize CT as it unfolds during such practices.

Teachers play a pivotal role in mediating STEM learning in early childhood classrooms. Their ability to notice and interpret CT within subject learning activities shapes how CT is supported in everyday teaching. However, CT is often described as abstract and difficult to observe, particularly in non-coding environments. This highlights the need to examine how CT becomes visible in classroom activities and how teachers come to recognize it as part of children's learning processes. This exploratory qualitative study investigates preschool teachers' awareness of computational thinking learning experiences as children created stop-motion animations within a STEM curriculum. Two animation projects, one centered on children's awareness of campus cultural architecture and the other on a soccer-themed story, were implemented within the same instructional context. By analyzing teachers' reflections on children's animation-making processes, this study aims to illuminate how CT becomes recognizable as a visible, experiential, and design-oriented form of thinking within early childhood STEM learning.

2. Literature Review

2.1. Computational Thinking in Early Childhood Education

Computational thinking (CT) has been widely recognized as a key component of early childhood education, encompassing practices such as sequencing, decomposition, pattern recognition, abstraction, and algorithmic reasoning (Wing, 2006). In early childhood contexts, CT is understood not as formal programming but as a way of thinking that develops through play-based, hands-on, and representational activities embedded in classroom experiences (Nouri et al., 2020). Scholars emphasize introducing CT to children aged 3–6 through developmentally appropriate learning opportunities (Sullivan & Bers, 2016). Research shows that young children can engage in CT-related practices when activities are grounded in meaningful contexts (Bers, 2018). Through interactions with materials, narratives, and visual representations, children organize sequences, break down tasks, recognize patterns, and construct simple procedures. These processes highlight the relationship among CT development, embodied action, narrative construction, and creative expression. Recent studies further indicate that creative and art-based activities can provide developmentally appropriate contexts for CT engagement. For example, Leung (2024) found that animation art activities grounded in constructionist approaches supported young children's conceptualization of CT. When CT is embedded within subject learning, such as storytelling and design tasks, it becomes part of children's everyday learning experiences (Strawhacker & Bers, 2019). This underscores the need to examine how CT is enacted and recognized in classroom practice.

2.2. Stop-Motion Animation as a Context for Computational Thinking

Research indicates that young children can meaningfully engage with digital tools in creative activities such as digital book creation, filmmaking, music production, and photography (Flewitt et al., 2015; Dezuanni et al., 2015). These activities require planning, representation, and revision processes aligned with foundational CT practices. Kotsopoulos et al. (2017) conceptualize CT learning as progressing through unplugged activities, tinkering, making, and remixing, with making-oriented experiences supporting CT through hands-on design and iterative refinement. Stop-motion animation, as a form of digital making, integrates physical manipulation, visual representation, and sequential design, offering a developmentally appropriate context for CT in early childhood settings. Building on this perspective, the present study examines how children's engagement in stop-motion animation supports problem solving, design planning, and iterative revision within classroom practice, contributing to understanding how CT can be enacted and observed through creative digital production.

2.3. Teachers' Awareness of Computational Thinking in Classroom Practices

Teachers play a critical role in supporting and interpreting CT within classroom activities. Their ability to recognize CT influences how learning experiences are scaffolded and sustained. Prior research highlights challenges teachers face in identifying CT, particularly when it is embedded within non-coding or interdisciplinary activities (Yadav et al., 2016; Hsu et al., 2018). More recent research suggests that teachers vary in their perceptions of the value of CT and often encounter difficulties integrating it meaningfully into instruction (Feng & Yang, 2022). Although observing learners' interactions with materials and representations may support teachers' understanding of CT as a process-oriented form of thinking (Voogt et al., 2015), limited research has examined teachers' awareness of CT in early childhood STEM classrooms, particularly in creative, human-centered contexts such as stop-motion animation. Addressing this gap, the present study investigates how preschool teachers recognize CT learning experiences through children's engagement in stop-motion animation within a STEM curriculum.

3. Methods

3.1. Research Design

This study employed an exploratory qualitative design to examine preschool teachers' awareness of children's computational thinking (CT) learning experiences during STEM-integrated stop-motion animation activities. The focus was on teachers' noticing and interpretation of CT-related practices as children planned, filmed, and revised animations.

3.2. Participants and Context

Participants were three preschool teachers from a public elementary school–affiliated preschool in Taiwan. Teacher A (15 years), Teacher B (12 years), and Teacher C (10 years) had an average of 12.3 years of teaching experience. All teachers had prior experience in STEAM-integrated teaching and had continuously participated in an innovative teaching project emphasizing computational thinking. The preschool, located in an urban Indigenous education school designated by the Ministry of Education, integrates culturally responsive practices and Indigenous knowledge into daily learning.

3.3. Stop-Motion Animation Activities

Within the integrated STEM curriculum, two stop-motion animation projects were implemented: Guāla Guāla House, which emerged from children's lived experiences with campus cultural architecture, and Soccer Match, which documented children's participation in an interschool football event. While the themes differed, both projects followed the same instructional structure and animation-making sequence (Figure 1). Across both projects, teachers guided children through a structured process that included recalling and organizing experiences, developing storyboards and preparing materials, capturing frame-by-frame movements, and reviewing and revising animations through playback (Figure 2). Through this process, children translated lived experiences into sequenced visual representations and progressively refined their work based on collective discussion and visual feedback. The animation-making activities emphasized iterative design, temporal sequencing, and representational decision-making. These processes were embedded within regular classroom practices across learning areas and were implemented using developmentally appropriate strategies aligned with early childhood STEM education.

Figure 1. Stop-motion setup for the Guāla Guāla House project.

Figure 2. Teacher-guided frame-by-frame animation filming.

3.4. Data Collection

Data were collected from two sources: semi-structured post-implementation interviews with the three teachers, and classroom instructional materials, including curriculum documentation, storyboards, production records, and completed animations. Interview data focused on teachers' reflections on children's learning processes and CT practices observed during animation-making. Instructional materials were used to contextualize and triangulate teachers' accounts.

Figure 3. Children's storyboard for the Soccer Match stop-motion animation.

Figure 4. Children producing and reviewing the Soccer Match stop-motion animation.

3.5. Data Analysis

Interview data were transcribed verbatim, and instructional materials were organized chronologically. Two researchers conducted qualitative coding in three stages. First, open coding identified segments related to children's actions, decision-making, and teachers' interpretations. Second, codes were compared and grouped into broader categories reflecting recurring patterns. Third, emergent categories were aligned with established CT dimensions, including decomposition, abstraction, sequencing, algorithmic thinking, pattern recognition, debugging, and human-in-the-loop control. Coding was iterative, and discrepancies were resolved through discussion until consensus was reached. To enhance the credibility of the analysis, data from interviews and instructional materials were triangulated, and interpretations were refined through ongoing discussion between the two researchers.

3.6. Ethical Considerations

All teachers provided informed consent. Parental consent was obtained for the use of classroom documentation and visual materials. Pseudonyms were used, and all data were handled in accordance with ethical standards for educational research.

4. Findings

4.1. Stop-Motion as a Catalyst for Decomposition and Narrative Structuring

All three teachers described how animation required children to reorganize lived experiences into structured representations. Teacher A explained that her goal was to help children "use life experiences to record animation" and "learn new technological skills while increasing confidence." Teacher B recalled that the project began with children's discussion about flooding after a typhoon, where they proposed imaginative solutions such as "building a drainage system" or "letting an elephant suck the water." Through collective discussion, children organized events into coherent sequences. Teacher B noted, "We sorted out the whole process into a timeline," while Teacher C stated, "When children mentioned important events, I recorded them and confirmed the plot." Teachers observed that children gradually broke experiences into characters, settings, actions, and temporal order. As Teacher A reflected, "We slowly broke it down into main characters, settings, and plot." In the Soccer Match project, decomposition became more independent. Teacher A stated, "The older children completed everything independently because it was their own story." Children organized their experience around the theme of "courage," demonstrating abstraction of a central idea from a complex event.

4.2. Frame-by-Frame Filming as Sequencing and Algorithmic Reasoning

Teachers consistently described filming as the point where CT became most visible. During the typhoon scene, Teacher A observed, "Each picture couldn't move too much; otherwise, it wouldn't look real." Teacher B noted that children learned to "move a little and take more photos. More photos are better than fewer." Teacher C added, "The house should shake left and right to look like a typhoon." Teachers also highlighted procedural organization. Teacher B explained, "One child pressed the tablet, and one person moved each character. That worked best." In the second project, children demonstrated increased awareness of production steps, recognizing the need for "scripts, characters, background, props, music," and asking about "the next step." Teacher C observed that children managed pacing more effectively, reflecting growing control of temporal sequencing.

4.3. Iteration, Debugging, and Refinement through Playback

Playback and revision were repeatedly identified as moments when CT became visible. Teacher B reported that children noticed characters "floating" or overlapping and repositioned them accordingly. Teacher C described how children felt the first version ended "too quickly" and suggested extending the storm scene: "The wind can blow longer, and the rain can fall more." After adding frames, children perceived the animation as more complete. Teacher B noted that children recognized the relationship between photo quantity and smoothness: "One scene needs about 20–30 photos." Voice recording also became a site of refinement. Teacher A explained, "We listened again and marked emotions next to each sentence," and Teacher C described how children adjusted tone to convey tension. In the second project, children monitored continuity across filming days; as Teacher B stated, "They used tape to fix the background position so it wouldn't shift."

4.4. Human-in-the-Loop Control

Teachers emphasized that animation required continuous human decision-making. Teacher C reflected, "Stop-motion makes it easier to observe children's thinking," noting that children's ideas became "more detailed" in the second project. Teacher B similarly observed increased proactivity in anticipating workflow and solving problems. The frame-by-frame nature of animation externalized children's decisions and made planning, sequencing, and revision observable within classroom practice.

5. Discussions

This study demonstrates how stop-motion animation, situated within an integrated STEM curriculum, provided a context in which children's computational thinking became visible to teachers. Across the Guāla Guāla House and Soccer Match projects, CT was enacted through children's narrative organization, frame-by-frame control, and iterative revision. Practices such as decomposition, sequencing, abstraction, algorithmic reasoning, and debugging emerged as children translated lived experiences into structured storyboards and refined animations through playback and discussion. These processes enabled teachers to recognize CT not as isolated technical skills, but as thinking embedded within creative and representational classroom activity. Importantly, this study focuses on teachers' awareness of computational thinking as interpreted through classroom practice, with an emphasis on how teachers notice and make sense of children's learning processes in context. The findings also highlight the importance of human-in-the-loop control. Stop-motion animation required continuous human judgment in coordinating movement, sequencing actions, maintaining continuity, and adjusting expressive elements. Teachers' awareness of CT was closely linked to these moments of decision-making and reflection. In this way, CT appeared as a human-centered, iterative design process shaped by collaboration and shared problem-solving. Pedagogically, the study suggests that creative digital production within STEM can support teachers' recognition of CT by externalizing children's thinking. Because stop-motion animation makes planning, sequencing, and revision visible, it offers early childhood educators a developmentally appropriate means of observing and articulating CT without relying on formal coding instruction.

This study is limited by its small sample size and specific context, as all teachers were from the same school and had prior engagement in a STEAM initiative. The findings should therefore be interpreted within this context. Future research may examine teachers' awareness across broader contexts and incorporate additional data sources, such as direct observations, video analysis, or children's verbal data, to further explore CT learning processes.

References

Feng, S., & Yang, D. (2022). Teachers' perceived value, challenges, and advice for implementing computational thinking in elementary classrooms. *Journal of Technology and Teacher Education*, *30*(3), 293-320. https://doi.org/10.70725/048319epkkuu

Hsu, T.-C., Chang, S.-C., & Hung, Y.-T. (2018). How to learn and how to teach computational thinking: Suggestions based on a review of the literature. *Computers & Education*, *126*, 296–310. https://doi.org/10.1016/j.compedu.2018.07.004

Kotsopoulos, D., Floyd, L., Khan, S., Namukasa, I. K., Somanath, S., Weber, J., & Yiu, C. (2017). A pedagogical framework for computational thinking. *Digital Experiences in Mathematics Education*, *3*(2), 154–171. https://doi.org/10.1007/s40751-017-0031-2

Leung, S. K. Y., Wu, J., Li, J. W., Lam, Y., & Ng, O.-L. (2024). Examining young children's computational thinking through animation art. *Early Childhood Education Journal*, *53*(5), 1563–1575. https://doi.org/10.1007/s10643-024-01694-w

Nouri, J., Zhang, L., Mannila, L., & Norén, E. (2020). Development of computational thinking, digital competence and 21st century skills when learning programming in K-9. *Education Inquiry*, *11*(1), 1–17. https://doi.org/10.1080/20004508.2019.1627844

Strawhacker, A., & Bers, M. U. (2019). What they learn when they learn coding: Investigating cognitive domains and computer programming knowledge in young children. *Educational Technology Research and Development*, *67*(3), 541–575. https://doi.org/10.1007/s11423-018-9622-x

Sullivan, A., & Bers, M. U. (2016). Robotics in the early childhood classroom: Learning outcomes from an 8-week robotics curriculum in pre-kindergarten through second grade. *International Journal of Technology and Design Education*, *26*(1), 3–20. https://doi.org/10.1007/s10798-015-9304-5

Wing, J. M. (2006). Computational thinking. *Communications of the ACM*, *49*(3), 33–35. https://doi.org/10.1145/1118178.1118215

Yadav, A., Gretter, S., Good, J., & McLean, T. (2016). Computational thinking for teacher education. *Communications of the ACM*, *59*(4), 55–62. https://doi.org/10.1145/2994591

Hsu, T. C., & Ogata, H. (Eds.). (2026). *Proceedings of the 10th International Conference on Artificial Intelligence, Computational Thinking and STEM Education* .Kumamoto: The Asia-Pacific Society for Computers in Education.

Supporting Computational Thinking in Video-Based Mathematics with AI-Mediated Reflection

Leo Andrei CRISOLOGO[1*], Ching-Huei CHEN[2] & Yu-Hui CHANG[3]
[1] International Graduate Program of Education and Human Development,
[2,3] Institute of Education and Center for Teacher Education, National Sun Yat-Sen University, Taiwan
*crisologoleoandrei@gmail.com

Abstract: *Computational thinking is central to mathematical problem solving, yet many of its processes remain implicit during instruction. In video-based mathematics learning, students can follow demonstrated procedures without articulating how ideas connect or how errors are resolved. Generative AI tools introduce new possibilities for supporting independent learning, but many systems provide full solutions that risk reducing engagement in decomposition, evaluation, and revision. This study investigates whether AI can be designed to scaffold computational thinking without replacing student reasoning.*

We developed an AI-mediated reflection system for secondary calculus lessons on solids of revolution. The system provides brief correctness feedback followed by a single targeted reflective prompt, without providing complete solutions and step-by-step procedures. Twenty-five students from an urban science high school in the Philippines participated in a two-week video-based instructional sequence. Students were assigned to either a fixed written reflection condition (n = 13) or an AI-mediated reflection condition (n = 12). Reflection episodes served as the unit of analysis and were coded for four computational thinking constructs: decomposition, algorithmic reasoning, evaluation, and debugging. Epistemic Network Analysis was used to compare how these constructs were structurally connected across conditions. Primary findings showed that computational thinking appeared in both conditions, but its organization differed. Fixed written reflection conditions emphasized procedural explanation, with strong connections between decomposition and algorithmic reasoning. In contrast, AI-mediated conditions showed stronger links between decomposition and abstraction, suggesting greater attention to structural relationships such as variable roles and axis orientation. Evaluation in the fixed condition was more closely tied to verifying steps, while in the AI condition it was more connected to examining structural coherence. Debugging was infrequent overall but
more explicit in AI-mediated sessions.

The findings indicate that AI-mediated reflection can reorganize how computational thinking processes are coordinated during learning. When designed to withhold solutions and prompt explanation, generative AI can function as reflective scaffolding that supports structural reasoning while maintaining student responsibility for problem solving.

Keywords: Computational thinking, Generative AI, AI-mediated reflection, Epistemic Network Analysis

1. Introduction

Computational thinking (CT) is central to mathematics education (Wing, 2006). In mathematics, CT includes decomposition, abstraction, algorithmic reasoning, evaluation, and debugging (Li et al., 2020; Shute et al., 2017). Although essential to problem solving, these processes often remain implicit when students follow demonstrated procedures without articulating underlying structure (Lodi & Martini, 2021).

Video-based instruction is common in secondary mathematics (Awang et al., 2025; Hwang & Tu, 2021). While it supports flexible pacing, it shifts responsibility for monitoring and evaluation onto the learner (Engelbrecht & Borba, 2024). Without structured reflection, students may reproduce procedures without examining structural coherence.

Generative AI introduces new possibilities for supporting independent reasoning (Debets et al., 2024; Kökver et al., 2025). However, many AI tools provide full solutions or procedural steps (Su & Yang, 2023), potentially reducing engagement in CT processes. The question is whether AI can scaffold computational thinking rather than replace it.

This study examines how generative AI can be structured to support CT during video-based calculus instruction. The research question guiding this study is: How does AI-mediated reflection influence the expression and structural organization of computational thinking during video-based mathematics learning? We compare AI-mediated reflection with fixed written prompts using coded reflection episodes and Epistemic Network Analysis (ENA) to examine not only whether CT processes appear, but how they are connected.

2. Computational thinking and reflection

Computational thinking refers to formulating problems in ways that enable structured solution procedures (Duckworth & Fraillon, 2025). Its core components of decomposition, abstraction, algorithmic reasoning, and debugging extend beyond computer science and align with mathematical problem solving (Shute et al., 2017). Reflection provides a context in which these processes become visible. When students articulate reasoning, justify steps, or reconsider errors, they externalize CT processes that might otherwise remain implicit (Demirel et al., 2015). Reflection episodes therefore offer a means to examine both the presence and structural coordination of CT during learning.

3. Design of the AI-mediated reflection system

The AI system in this study was designed to scaffold CT without providing solutions. It withheld step-by-step procedures and final answers, offering only correctness feedback followed by a single targeted reflective prompt. Prompts directed students to articulate procedural structure, examine variable relationships, evaluate coherence, and reconsider possible weaknesses in their reasoning. By maintaining student responsibility for explanation, checking, and revision, the system aimed to integrate decomposition, algorithmic articulation, evaluation, and debugging into the reflection process.

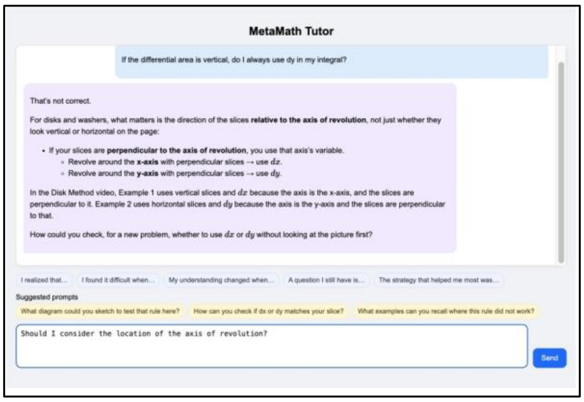


Figure 1. User interface of the AI-mediated reflection system

4. Context and methods

4.1. Participants and procedure

The study was conducted over a two-week instructional sequence on solids of revolution. Students engaged with a series of video-based lessons individually at home designed to introduce and explain key concepts and procedures related to the lesson. The participants were twenty-five secondary students from an urban science high school in the Philippines. All instructional materials, including videos and assessments, were consistent across participants.

For each lesson, students first watched the instructional video, which introduced key concepts and demonstrated worked examples. After viewing the lesson, they were required to submit a reflection describing their understanding and reasoning. In the control condition, this reflection was written in response to fixed prompts. In the experimental condition, students entered their reflection into a chatbot interface that function by the AI system design. After each student message, the chatbot provided brief correctness feedback followed by a single targeted prompt for clarification or revision. This exchange could continue for multiple turns within the same session. The AI system was used only during the reflection phase and, although it may provide brief conceptual clarifications, it did not give correct answers, worked solutions, or step-by-step procedures.

Students were assigned to one of two conditions: In the control condition (n=13), learners completed fixed written reflection prompts after each lesson. These prompts required students to describe their understanding and consider how they might improve their approach. In the experimental condition (n=12), learners completed their reflections through interaction with the AI chatbot. Both groups received the same instructional content and engaged with the same video lessons. The only difference between conditions was the format and structure of the reflection activity.

4.2. Data sources

Data were collected from multiple sources to examine how computational thinking processes were expressed and organized during learning. These included written reflections from the control group, chatbot interaction transcripts from the experimental group, post-lesson interview data, and interaction timing logs from the AI system.

4.3. Coding and data analysis

The primary unit of analysis was the reflection episode. For each lesson, a reflection episode consisted of either one complete written reflection (control condition) or one full chatbot session (experimental condition), including all student-AI exchanges within that session. All CT codes identified within the same episode were combined at the episode level.

Reflection episodes were coded for four computational thinking constructs: decomposition, algorithmic reasoning, evaluation, and debugging. Each construct was coded as either present or absent within a reflection episode.

Epistemic Network Analysis (ENA) was used to model co-occurrence of CT constructs within episodes. In the ENA model, nodes represented CT constructs and edges represented their co-occurrence within an episode, allowing comparison of structural organization across conditions.

5. Results

5.1. Expression of CT processes

CT appeared in both conditions but differed in emphasis. In the fixed written reflection condition, reflections focused on structural explanation. For example, one participant reported that *"getting the area of a solid of revolution involves breaking it down into infinite slices" (*F03, written reflection). Here, decomposition and abstraction are linked through explanation of structural roles.

In the AI-mediated reflection condition, reflections more frequently used ordered procedural language and explicit correction. For example: *"I could check first if I am using the correct differential (dy or dx)"(*A02, reflection episode). Debugging statements such as correcting bounds or differentials appeared more often in AI-mediated sessions.

Overall, based on the coding of reflection episodes, participants in fixed written reflections inclined to emphasize conceptual structure, while participants in AI-mediated reflection sessions showed clearer procedural articulation and explicit error recognition.

5.2. Connections of CT engagements

Epistemic Network Analysis was used to examine how CT constructs were connected in the fixed written reflection and AI-mediated reflection conditions. Units were defined at the student level. Because each conversation was coded as one line, a stanza window of 1 was used so that all CT codes within the same episode were treated as co-occurring.

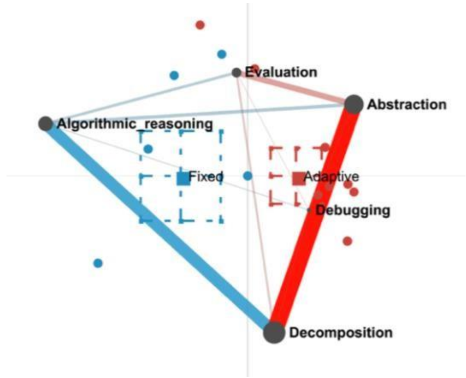


Figure 2. Epistemic network comparison of computational thinking constructs in AI-mediated reflection and fixed written reflection conditions. Nodes represent CT constructs, and edge thickness reflects co-occurrence strength within reflection episodes. Red indicates AI-mediated reflection; blue indicates fixed written reflections.

The network graphs show a clear difference between conditions. The centroids for AI-mediated reflection and fixed written reflection episodes are separated along the main projection axis. This means the two conditions differ not only in how often CT processes appear, but also in how those processes are connected.

In the fixed written reflection condition, the strongest connection is between decomposition and algorithmic reasoning. When students break a problem into parts, they tend to link this to ordered steps. In this condition, CT is organized around procedural explanation: break the problem down, then describe the steps to solve it.

In the AI-mediated reflection condition, the strongest connection is between decomposition and abstraction. When students break the problem into parts, they are more likely to connect this to identifying structural relationships, such as the role of variables, axis orientation, or inner and outer radii. Abstraction is more central in the AI-mediated reflection network, meaning structural reasoning plays a larger role in these sessions.

Evaluation also differs across conditions. In the fixed written reflection network, evaluation connects more strongly to algorithmic reasoning. In the AI-mediated reflection network, Evaluation connects more to Abstraction. This suggests that in the fixed written reflection condition, checking is tied to verifying steps while in the AI-mediated reflection condition it is connected to whether the structure of the solution makes sense.

Debugging appears rarely in both conditions and is not central in either network. This is consistent with the strict rule used to code debugging only when students clearly identified and corrected an error.

Overall, the results show that the two conditions organize computational thinking differently. Fixed written reflections show stronger procedural connections. AI-mediated reflection sessions show stronger links between breaking the problem apart and identifying its structure. These findings indicate that AI-mediated dialogue may support more structurally focused problem representation.

6. Discussion and conclusion

Both conditions engaged core CT processes but organized them differently. In the fixed written condition, decomposition was most strongly connected to algorithmic reasoning, suggesting a procedural orientation. In the AI-mediated reflection condition, decomposition was more strongly linked to abstraction, indicating greater attention to structural relationships. Evaluation also differed: in fixed written reflections it aligned with step verification, whereas in AI-mediated reflection sessions it aligned with structural coherence. Debugging was infrequent overall but more explicit in AI-mediated sessions.

These results show that AI-mediated reflection does not merely increase CT frequency but can reorganize how CT components are coordinated. Generative AI systems can be designed to function as reflective scaffolding (Chen, Chen, & Lai, 2025) that supports structural reasoning while maintaining student responsibility for problem solving. Future studies could further examine the reflective design of scaffolding within the AI-mediated system.

References

Awang, L. A., Yusop, F. D., & Danaee, M. (2025). Current practices and future direction of artificial intelligence in mathematics education: A systematic review. *International Electronic Journal of Mathematics Education, 20*(2), em0823. https://doi.org/10.29333/iejme/16006

Chen, S. Y., Chen, W. C., & Lai, C. F. (2025). *Generative AI as a reflective scaffold in a UAV-based STEM project: A mixed-methods study on students' higher-order thinking and cognitive transformation.* Education and Information Technologies, 30, 24787–24814. https://doi.org/10.1007/s10639-025-13758-4

Debets, T., Banihashem, S. K., Joosten-Ten Brinke, D., Vos, T. E. J., Maillette de Buy Wenniger, G., & Camp, G. (2025). Chatbots in education: A systematic review of objectives, underlying technology and theory, evaluation criteria, and impacts.

Computers & Education, 234, 105323. https://doi.org/10.1016/j.compedu.2025.105323 Demirel, M., Derman, I., & Karagedik, E. (2015). A study on the relationship between

reflective thinking skills towards problem solving and attitudes towards mathematics. Procedia – *Social and Behavioral Sciences, 197*, 2086–2096. https://doi.org/10.1016/j.sbspro.2015.07.326

Duckworth, D., & Fraillon, J. (2025). *Computational thinking framework*. In J. Fraillon & M. Rožman (Eds.), *IEA International Computer and Information Literacy Study 2023*. Springer. https://doi.org/10.1007/978-3-031-61194-0_3

Engelbrecht, J., & Borba, M. C. (2024). Recent developments in using digital technology in mathematics education. ZDM–*Mathematics Education, 56*, 281–292. https://doi.org/10.1007/s11858-023-01530-2

Hwang, G.-J., & Tu, Y.-F. (2021). Roles and research trends of artificial intelligence in mathematics education: A bibliometric mapping analysis and systematic review. *Mathematics, 9*(6), 584. https://doi.org/10.3390/math9060584

Kökver, Y., Pektaş, H. M., & Ç elik, H. (2025). Artificial intelligence applications in education: Natural language processing in detecting misconceptions. *Education and Information Technologies, 30,* 3035–3066. https://doi.org/10.1007/s10639-024-12919-1

Li, Y., Schoenfeld, A. H., diSessa, A. A., Graesser, A. C., Benson, L. C., English, L. D., & Duschl, R. A. (2020). Computational thinking is more about thinking than computing. *Journal for STEM Education Research, 3*, 1–18. https://doi.org/10.1007/s41979-020-00030-2

Lodi, M., & Martini, S. (2021). Computational thinking, between Papert and Wing. *Science & Education, 30*(4), 883–908. https://doi.org/10.1007/s11191-021-00202-5

Shute, V. J., Sun, C., & Asbell-Clarke, J. (2017). Demystifying computational thinking. *Educational Research Review, 22*, 142–158. https://doi.org/10.1016/j.edurev.2017.09.003

Su, Y., & Yang, X. (2023). Current practices and future direction of artificial intelligence in mathematics education: A review. *International Electronic Journal of Mathematics Education, 18*(3), Article em0738. https://doi.org/10.29333/iejme/16006

Wing, J. M. (2006). Computational thinking. *Communications of the ACM, 49*(3), 33–35. https://doi.org/10.1145/1118178.1118215

Hsu, T. C., & Ogata, H. (Eds.). (2026). *Proceedings of the 10th International Conference on Artificial Intelligence, Computational Thinking and STEM Education* .Kumamoto: The Asia-Pacific Society for Computers in Education.

A PDCA-Based Instructional Framework for AI-Assisted Programming

YungYu ZHUANG[1*], Wei-Hsin MA[2], Fang-Tzu TSAI[3], Jen-Hang WANG[4] & Gwo-Dong CHEN[5]
[1,2,3,4,5] Dept. of Computer Science and Information Engineering, National Central University, Taiwan
[*]yungyu@ncu.edu.tw

Abstract: *Since now AI can be used to generate program code according to natural-language sentences, anyone can be a programmer to create programs to solve computational problems in their domains. However, for novice programmers, ensuring that AI-generated implementations meet their needs remains very challenging. Although programming has been recognized as an effective way to develop learners' computational thinking, learners who fully trust AI and accept generated code without thinking may fail in learning. To address this issue, we develop a comprehensive framework to guide learners in learning and using AI-assisted programming. In the proposed framework, we adopt the simple yet effective PDCA cycle and introduce the "plan-for-check" concept, which connects the Plan stage and the Check stage, in response to requirements analysis and software testing, respectively. With our framework, students can easily learn to practice key software engineering concepts in their AI-assisted programming. To evaluate our framework, we implemented a learning system to support the methodology and conducted an experiment comparing the learning performance achieved with our framework to that of directly chatting with AI. The experimental results show that our methodology can help learners acquire the ability to effectively use AI in programming.*

Keywords: Generative AI, programming education, computational thinking, PDCA cycle, AI-assisted programming

1. Introduction

As generative AI (GenAI) tools mature and become more popular, people can use AI to quickly generate programs from natural-language instructions rather than traditional coding. People may provide sentences in English or any language they are familiar with as prompts to GenAI tools like ChatGPT to generate concrete implementations in a specific programming language, such as Python. It means that everyone can be a programmer with AI's assistance, creating programs to solve computational problems in their domains.

However, ensuring the generated implementations fit programmers' needs is very challenging. To experienced programmers, it might not be tough since they know how to verify the generated programs, as they do for code review. On the other hand, for novice programmers, the question is how they should learn and use AI-assisted programming. Several research activities have highlighted these issues, including the tendency to accept AI's suggestions without checking and the lack of metacognitive awareness (Prather et al., 2023; Prather et al., 2024).

Even though programming has been recognized as an effective means of developing a conceptual model of computational thinking (CT) (Tikva & Tambouris, 2021), the emergence of GenAI tools complicates this picture. Without a comprehensive framework to guide students in learning and using AI-assisted programming, students may fully trust AI and accept generated code without thinking (Vaithilingam et al., 2022; Zhang et al., 2023), leading to overreliance on GenAI tools. As many research studies have highlighted, reading and learning from debugging are important for learning programming (Lowe, 2019; Marceau et al., 2011). Consequently, students can save time on debugging, but also miss the chance to learn from it.

To address this issue, this study was conducted to consider and answer the following two research questions (RQs):

(1) Can we introduce a simple strategy, such as the PDCA cycle, to develop a framework for guiding students to learn and use AI-assisted programming?

(2) Can such a framework be used to improve students' learning achievement, compared with using GenAI tools without it?

We developed an instructional framework for AI-assisted programming based on the PDCA (Plan-Do-Check-Act) cycle to guide learners to consider "plan-for-check" in their programming tasks. As scaffolding, it can help learners learn and use GenAI tools in programming. We then implemented a learning system and conducted an experiment to evaluate our methodology. Note that this paper reports only the instructional design with a preliminary result. For the metacognitive support in our framework, we are analyzing the data for further discussion.

2. Literature Review

Below, we review the literature on programming with GenAI to further clarify the issue we would like to address, and then discuss the application of the PDCA cycle in education, explaining why we introduce this strategy into AI-assisted programming. Finally, we discuss requirements analysis and software testing, which map to the Plan and Check stages in the PDCA cycle, respectively, to highlight the need of "plan-for-check" in our proposed methodology.

2.1. Programming with Generative AI

Nowadays, computer programming is no longer limited to computer science students; it is a common tool and skill for resolving issues in specialized domains. With GenAI tools such as ChatGPT and Gemini, learners can ask for solutions and concrete implementations through chat, then copy the generated programs for use. On the other hand, programmers may also use AI-assisted programming, such as Copilot, writing comments in an IDE (integrated development environment) to generate partial programs. Some practice, including Vibe Coding and several research results, already showed that GenAI can help learners in programming by lowering the learning threshold and reducing mental effort (Deng et al., 2025; Groothuijsen et al., 2024; Yilmaz & Karaoglan Yilmaz, 2023).

On the other hand, there are many research findings suggesting that how learners use GenAI in problem-solving is critical (Hou et al., 2026; Hou et al., 2025), and the risk of overreliance on GenAI has been revealed as well (Nathaniel et al., 2025; O'Brien, 2025; Sergeyuk et al., 2025; Zhou et al., 2025). Undoubtedly, GenAI has become an indispensable programming tool, but it also requires programmers to have a clear understanding of planning and testing.

2.2. PDCA Cycle in Education

The original concept of PDCA (Plan-Do-Check-Act) cycle was proposed by Shewhart and later modified and popularized by Deming (Moen & Norman, 2010). It is an iterative design for controlling the quality of processes or products in business. The PDCA cycle starts with the Plan stage, where a product plan is developed. After that, we implement the plan in the do stage and identify any issues within the product in the Check stage. Then, during the act stage, we adjust and refine the goal based on the identified issues to provide a better baseline for the next cycle. By repeating the PDCA cycle, the quality of products can be consistently controlled and improved.

The PDCA cycle has been applied in education, demonstrating its benefits for teaching practices (Huan & Nasri, 2022). It was also introduced into tangible programming education for children, reporting its effectiveness in developing students' structured programming skills, i.e., using sequence, repetition, and conditional branches, along with their reflective thinking (Gong et al., 2024). These research activities revealed the simplicity and potential of applying the PDCA cycle and motivated us to adopt it in our methodology.

2.3. Requirements Analysis and Software Testing

As a starting point, requirement analysis plays an essential role in software development. It is to analyze the conditions required to meet the product's needs, including considering system requirements, possible conflicts, documentation, and validation. Empirical evidence in software engineering shows that high-quality requirements analysis significantly influences downstream development activities, reducing rework and project failures (Frattini et al., 2023; Montgomery et al., 2022), while poorly managed requirements link to misunderstandings and such increased analysis effort and reduced project success rates (Tamai & Kamata, 2009).

On the other hand, software testing is a central technique used in software verification and validation to build confidence that an implementation meets its intended specifications (Adrion et al., 1982). Once software product development begins, the next question is how to determine whether the quality meets the requirements. Although software testing is such a critical aspect of the software development cycle, challenges persist in effectively teaching it (Fasolino et al., 2025). We integrated the concepts of requirements and software testing from software engineering and combined the simplified versions with the Plan and the Check stages in the PDCA cycle, respectively. Furthermore, we borrowed the idea of Test-Driven Development (TDD) (Janzen & Saiedian, 2005) and converted it into "plan-for-check" for the PDCA cycle in our methodology.

3. Methodology and Learning Design

In this section, we explain how the PDCA cycle was adopted in our AI-assisted programming framework, particularly the "plan-for-check" concept, and demonstrate the learning system we implemented for this framework.

3.1. The PDCA Cycle for AI-Assisted Programming

The learning design of our framework is shown in Figure 1, where each outer block represents a stage in the PDCA cycle: Plan, Do, Check, and Act. In each stage, we define what to do for programming with GenAI, respectively. In the Plan stage, learners need to carefully read the given problem description, analyze the requirements for the target program, and set up check items. During the requirement analysis, learners may also consult GenAI for advice. Unlike the formal way in software engineering, we focus on the input and output for novice programmers. In the Do stage, learners can ask AI to generate programs. Then, in the Check stage, learners must verify the generated programs with the check items set in the Plan stage. This is to ensure that the generated program meets the requirements and works well for learners. If any check fails, learners have to think about the reason and act on it during the Act stage, then return to the check items setting in the Plan stage. The goal of our framework is to help learners develop the ability to plan and check programs rather than relying entirely on GenAI. Just like existing automation tools, such as code generation from UML diagrams, GenAI should be a tool that generates implementations based on programmers' intentions, meaning programmers need to know how to determine whether the generated code meets their needs.

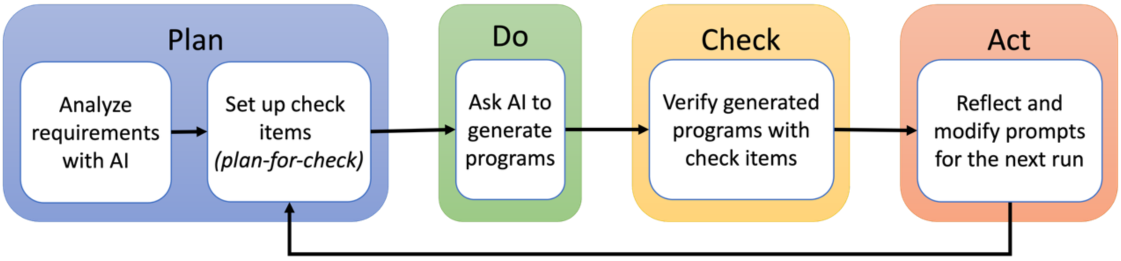


Figure 1. The PDCA cycle for AI-assisted programming in our framework.

3.2. The PDCA Cycle for AI-Assisted Programming

Inside the Plan stage, there is a key point: plan-for-check. In the proposed framework, we strictly ask learners to do planning based on how they will check. With this idea, learners can grasp what to plan and how to check. The idea of "plan-for-check" is borrowed from TDD in software development, where programmers should always consider input and output and prepare tests for their functions and modules. For novice programmers, we are not forcing them to apply formal testing; we are helping them understand how to consider verification at an early stage. It also helps learners imagine what the program should look like.

3.3. Our learning system

We developed a learning system based on our methodology as a scaffold to guide learners in programming with GenAI within the framework. Our learning system is a wizard-style, browser-based interface that integrates the OpenAI API to leverage the GPT-4o-mini model and connects to a MySQL database. The interface provides a combination of a panel to display the problem description, a form to set and list check items, an area to show generated code, and a console to display execution results, depending on the stage of the PDCA cycle the learner is in. Our learning system is designed to help follow and foster the ability to program with GenAI. After learners acquire the ability, they can program with any kind of GenAI without our learning system.

4. Experimental results and data analysis

To evaluate our methodology, we designed an experiment, as shown in Figure 2, to compare students' learning achievement when using our learning system with that when using GenAI tools directly. We recruited subjects from a university in Taiwan to participate in our experiments and randomly assigned them to the experimental and control groups. Note that students come from different departments and have less than one year of programming experience, excluding students from CS-related departments. The pedagogical activities every week consist of two parts: 10 minutes for learning material and 40 minutes for programming tasks. The learning material for both groups is delivered by the same instructor, with the same programming task, but during the 40 minutes of exercise, they use different systems to learn: the experimental group works on our learning system, while the control group chats directly with GenAI without the PDCA cycle or wizard. Regarding the design of programming tasks and the learning in the control group, we refer to the one designed by Vadaparty et al. (2024).

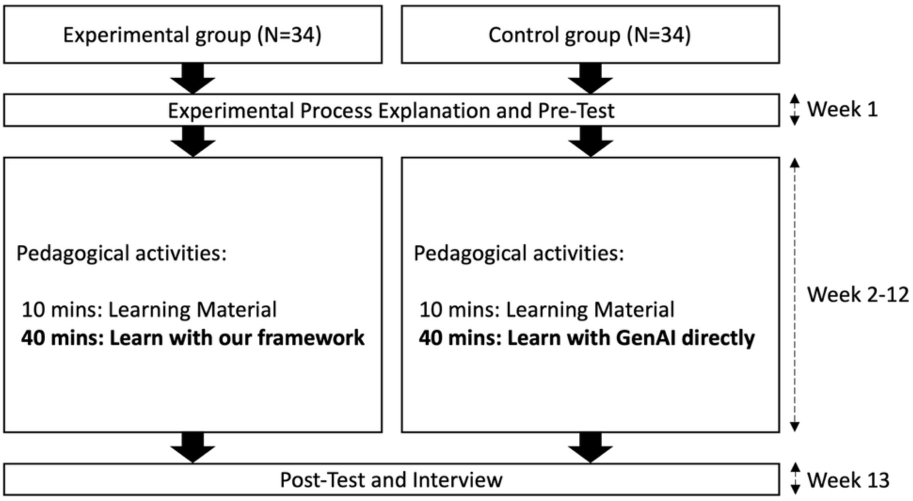


Figure 2. The experimental procedure.

For the tests, we prepared yet another programming task equivalent to those in weeks 2 to 12, focusing on application scenarios and requirements analysis. The goal is to design an ordering system for desserts that can handle customer orders and calculate payments for different item types, where customers' orders are sequences of items with arithmetic, and learners need to consider some exceptions and invalid cases. In this experiment, we took not only the program's results but also learners' modifications to prompts into account when defining the scoring rubric. The scoring is done by two experienced programming instructors and follows the standardized scoring procedure and consensus moderation.

After ensuring the samples were normally distributed using the Shapiro-Wilk test and confirming homogeneity of variances using Levene's test, the one-way analysis of covariance (ANCOVA) was employed to account for the difference in pre-test scores between the two groups. The ANCOVA results for learning achievement show that the adjusted mean post-test score of the experimental group is 62.145, while that of the control group is 53.931, indicating a significant difference between the two groups (F = 7.223, p = .009 < .01).

The numerical results indicate that the experimental group using our framework outperforms the control group. Thus, the answer to both RQs posed in the introduction is “yes”: we demonstrated the feasibility of designing a simple framework to guide learners and improve their learning outcomes. In interviews with subjects in the experimental group, several learners provided positive feedback, such as “learning the PDCA cycle helps me program with structural methods rather than just try-and-error” and “now I will proactively think about how to check the generated programs.” Although the analysis of our experiment reported here remains limited, this feedback also suggests that our framework is effective. To clarify our assumption about the metacognitive support provided by the framework, we are now analyzing experimental data and interview transcripts for further discussion.

5. Conclusions

To guide novice programmers seeking to benefit from GenAI, we developed a framework based on the PDCA cycle and introduced the “plan-for-check” concept. We also implemented a learning system based on it and conducted an experiment that demonstrated the feasibility of such a framework. In this paper, we present numerical results from our preliminary analysis. We are now analyzing the experimental results in detail to discuss the metacognitive support provided by our framework. How to further refine the Plan and Check stages for teaching is also include in our future work.

References

Adrion, W. R., Branstad, M. A., & Cherniavsky, J. C. (1982). Validation, Verification, and Testing of Computer Software. *ACM Comput. Surv.*, *14*(2), 159–192. https://doi.org/10.1145/356876.356879

Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2025). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. *Computers & Education*, *227*, 105224. https://doi.org/https://doi.org/10.1016/j.compedu.2024.105224

Fasolino, A. R., MarÍn, B., Vos, T. E. J., Mendes, A., Paiva, A. C. R., Cammaerts, F., Snoeck, M., Saadatmand, M., & Tramontana, P. (2025). Teachers’ Perspective on Software Testing Education. *ACM Trans. Comput. Educ.*, *26*(1), Article 10. https://doi.org/10.1145/3772090

Frattini, J., Montgomery, L., Fischbach, J., Mendez, D., Fucci, D., & Unterkalmsteiner, M. (2023). Requirements quality research: a harmonized theory, evaluation, and roadmap. *Requirements Engineering*, *28*(4), 507-520. https://doi.org/10.1007/s00766-023-00405-y Gong, X., Yu, S., Xu, J., Qiao, A., & Han, H. (2024). The effect of PDCA cycle strategy on pupils’ tangible programming skills and reflective thinking. *Education and information technologies*, *29*(5), 6383-6405. https://doi.org/10.1007/s10639-023-12037-4

Groothuijsen, S., van den Beemt, A., Remmers, J. C., & van Meeuwen, L. W. (2024). AI chatbots in programming education: Students’ use in a scientific computing course and consequences for learning. *Computers and Education: Artificial Intelligence*, *7*, 100290. https://doi.org/https://doi.org/10.1016/j.caeai.2024.100290

Hou, C., Zhu, G., Liu, Y., Sudarshan, V., Chong, J. L. L., Zhang, F. Y., Tan, M. Y. H., & Ong, Y. S. (2026). The effects of critical thinking intervention on reliance behaviors, problem-solving quality, and creativity during human-Generative AI collaborative learning. *Computers & Education*, *247*, 105576. https://doi.org/https://doi.org/10.1016/j.compedu.2026.105576

Hou, C., Zhu, G., Sudarshan, V., Lim, F. S., & Ong, Y. S. (2025). Measuring undergraduate students' reliance on Generative AI during problem-solving: Scale development and validation. *Computers & Education*, *234*, 105329. https://doi.org/https://doi.org/10.1016/j.compedu.2025.105329

Huan, C. W., & Nasri, N. M. (2022). Teacher teaching practices based on the PDCA model: A systematic literature review. *International Journal of Academic Research in Progressive Education and Development*, *11*(3), 542-553.

Janzen, D., & Saiedian, H. (2005). Test-driven development concepts, taxonomy, and future direction. *Computer*, *38*(9), 43-50. https://doi.org/10.1109/MC.2005.314

Lowe, T. (2019, 16-19 Oct. 2019). Debugging: The key to unlocking the mind of a novice programmer? 2019 IEEE Frontiers in Education Conference (FIE),

Marceau, G., Fisler, K., & Krishnamurthi, S. (2011). *Measuring the effectiveness of error messages designed for novice programmers* Proceedings of the 42nd ACM technical symposium on Computer science education, Dallas, TX, USA. https://doi.org/10.1145/1953163.1953308

Moen, R. D., & Norman, C. L. (2010). Circling back. *Quality Progress*, *43*(11), 22.

Montgomery, L., Fucci, D., Bouraffa, A., Scholz, L., & Maalej, W. (2022). Empirical research on requirements quality: a systematic mapping study. *Requirements Engineering*, *27*(2), 183- 209. https://doi.org/10.1007/s00766-021-00367-z

Nathaniel, J., Oyelere, S. S., Suhonen, J., & Tedre, M. (2025). Investigating the impact of generative AI integration on the sustenance of higher-order thinking skills and understanding of programming logic. *Computers and Education: Artificial Intelligence*, *9*, 100460. https://doi.org/https://doi.org/10.1016/j.caeai.2025.100460

O'Brien, G. (2025). Threats to scientific software from over-reliance on AI code assistants. *Nature Computational Science*, *5*(9), 701-703.

Prather, J., Reeves, B. N., Denny, P., Becker, B. A., Leinonen, J., Luxton-Reilly, A., Powell, G., Finnie-Ansley, J., & Santos, E. A. (2023). "It's Weird That it Knows What I Want": Usability and Interactions with Copilot for Novice Programmers. *ACM Trans. Comput.-Hum. Interact.*, *31*(1), Article 4. https://doi.org/10.1145/3617367

Prather, J., Reeves, B. N., Leinonen, J., MacNeil, S., Randrianasolo, A. S., Becker, B. A., Kimmel, B., Wright, J., & Briggs, B. (2024). *The Widening Gap: The Benefits and Harms of Generative AI for Novice Programmers* Proceedings of the 2024 ACM Conference on International Computing Education Research - Volume 1, Melbourne, VIC, Australia. https://doi.org/10.1145/3632620.3671116

Sergeyuk, A., Golubev, Y., Bryksin, T., & Ahmed, I. (2025). Using AI-based coding assistants in practice: State of affairs, perceptions, and ways forward. *Information and Software Technology*, *178*, 107610. https://doi.org/https://doi.org/10.1016/j.infsof.2024.107610

Tamai, T., & Kamata, M. I. (2009). Impact of Requirements Quality on Project Success or Failure. In

K. Lyytinen, P. Loucopoulos, J. Mylopoulos, & B. Robinson, *Design Requirements Engineering: A Ten-Year Perspective* Berlin, Heidelberg.

Vadaparty, A., Zingaro, D., IV, D. H. S., Padala, M., Alvarado, C., Benario, J. G., & Porter, L. (2024). *CS1-LLM: Integrating LLMs into CS1 Instruction* Proceedings of the 2024 on Innovation and Technology in Computer Science Education V. 1, Milan, Italy. https://doi.org/10.1145/3649217.3653584

Yilmaz, R., & Karaoglan Yilmaz, F. G. (2023). The effect of generative artificial intelligence (AI)-based tool use on students' computational thinking skills, programming self-efficacy and motivation. *Computers and Education: Artificial Intelligence*, *4*, 100147. https://doi.org/https://doi.org/10.1016/j.caeai.2023.100147

Zhou, X., Liang, P., Zhang, B., Li, Z., Ahmad, A., Shahin, M., & Waseem, M. (2025). Exploring the problems, their causes and solutions of AI pair programming: A study on GitHub and Stack Overflow. *Journal of Systems and Software*, *219*, 112204. https://doi.org/https://doi.org/10.1016/j.jss.2024.112204

A Smartphone-Enabled Progressive Web Application for Code-Free Introductory AI Education: Design and Classroom Evaluation

Shichao MA[1], Weizhong ZHENG[2], Emily Nicola STABLES[3], Victoria Anna YEO[4],
Mun Kay HO[5], Joshua W. K. HOb,[6*]
[1] Department of Computer Science, Hong Kong Baptist University, Hong Kong SAR, China
[2,3,4,5,6]School of Biomedical Sciences, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, China
[6]School of Biomedical Engineering, The University of Hong Kong, Hong Kong SAR, China
[*] jwkho@hku.hk

Abstract: *Accessible hands-on instruction in artificial intelligence remains difficult to implement in many K-12 and interdisciplinary settings because practical activities often require coding experience, managed computer laboratories, and substantial setup time. This paper presents a smartphone-enabled progressive web application for introductory machine learning education and reports a classroom-based evaluation with novice learners to provide preliminary evidence of its feasibility and educational value. The technical contribution is a browser-native instructional architecture that combines a pre-trained convolutional neural network, on-device fine-tuning, and a simplified mobile user interface so that learners can complete the supervised learning workflow (data collection, labeling, training, testing, and inference) without programming or specialized software installation. Educational effectiveness was examined using a one-group pre-test/post-test design with first-year non-CS university students (matched sample $n=28$). Paired conceptual quiz scores increased from $M=2.89$ ($SD=1.47$) to $M=3.71$ ($SD=1.38$), $t(27)=2.21$, $p=0.035$, with a mean gain of 0.82 points (95% CI [0.06, 1.58]). Post-workshop ratings indicated favorable perceived ease of use, usefulness, and self-efficacy/attitude ($n=33$; all construct means >4.0/5). The findings suggest that a comparatively lightweight PWA implementation can support scalable, low-cost, code-free AI learning experiences while preserving core conceptual learning outcomes.*

Keywords: AI education, educational technology, progressive web application, interdisciplinary education, computational thinking education, transfer learning

1. Introduction

Teaching core concepts in artificial intelligence (AI) and machine learning (ML) often depends on programming exercises, desktop computers, and software environments that are difficult to deploy in K-12 and interdisciplinary classrooms. This is a practical barrier in health-related programs, where many students are interested in AI applications but have limited prior coding experience. In addition to technical complexity, hardware provisioning and classroom logistics can make hands-on activities expensive and difficult to scale (AI Index Steering Committee, 2022).

To address this gap, we designed a smartphone-enabled progressive web application (PWA) that allows students to experience a complete supervised learning workflow on their own devices without coding. Unlike many introductory activities that depend on preconfigured desktops or notebooks, the present design shifts the delivery burden to a browser-native application that can be accessed through a QR code. This design is intended to reduce setup overhead for educators while retaining a genuine hands-on learning sequence rather than a passive demonstration.

In this paper, we report both the technical design and an empirical classroom evaluation involving first-year non-CS university students. The contribution is not a novel ML algorithm; rather, it is a deployable educational engineering pattern:

an easy-to-develop, easy-to-deploy PWA that packages transfer learning and mobile interaction into a classroom-ready tool for AI education. We further provide preliminary evidence from a pre-test/post-test study examining conceptual learning gains and student perceptions of usability, usefulness, and confidence. Pedagogically, the workshop is framed as a guided-discovery and experiential learning activity in which students manipulate data, observe model behavior, and infer core ML concepts from immediate feedback.

2. System and Workshop Design

2.1. Smartphone PWA for on-device ML learning

The application is a browser-accessible PWA for smartphones and tablets. Learners scan a QR code and use a simple mobile interface to add labeled examples, train a model, estimate accuracy, and classify new images. The workshop task uses a binary chest X-ray (CXR) pneumonia-versus-normal example, which is concrete for medical AI discussion and compatible with prior medical imaging work (Kermany et al., 2018).

To make training feasible in a short class, the app uses a pre-trained convolutional neural network and lightweight on-device fine-tuning rather than training from scratch. This transfer learning design reduces the number of new examples required while preserving the pedagogical value of data collection, labeling, training, and testing (Pan & Yang, 2010). The interface is intentionally task-driven, exposing only a small set of actions (add data, train, update accuracy, classify, reset) so students can focus on concepts instead of programming syntax.

2.2. Technical implementation and transferability

The prototype converts an ML teaching activity into a browser-deliverable tool by adapting a pre-trained convolutional neural network to a binary CXR task and converting the model from a Python/Keras workflow for browser execution using TensorFlow.js tooling (Smilkov et al., 2019).

The front-end PWA lowers setup complexity and can be reused across topics by changing labels, preloaded models, and media assets. Compared with notebook- or IDE-based exercises, it reduces technical overhead for both instructors and learners.

Figure 1 summarizes the overall technical and pedagogical design, including browser-side deployment, model preparation, and smartphone-based data capture/labeling.

2.3. Workshop learning flow

The workshop is a guided hands-on activity in which students use the PWA to collect labeled CXR examples, train a personalized model on-device, estimate performance on a built-in withheld test set, and test inference on new images.

The activity connects app actions to ML concepts (labeled data, train/test split, transfer learning, data quality). Students are encouraged to experiment with less data, poor-quality images, and incorrect labels so that model behavior becomes a basis for inference and discussion.

The workshop is intentionally heuristic: participants explore, test, and infer ML concepts from the system's responses rather than receiving step-by-step explicit teaching of each quiz concept. This guided-discovery structure is intended to support experiential concept building: learners iteratively collect data, inspect outcomes, and reflect on how labeling, data quality, and testing affect model performance.

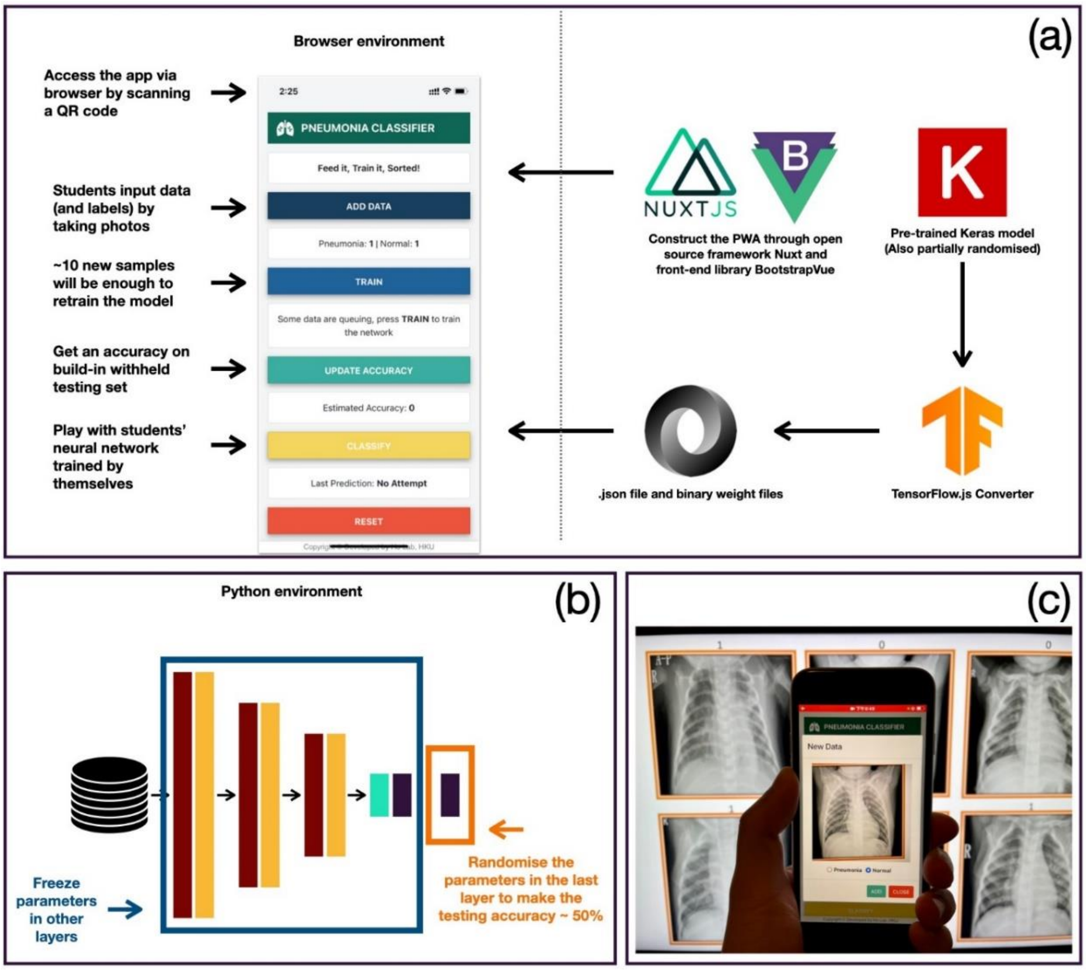


Figure 1. Technical and pedagogical overview of the PWA workshop. (a) Browser-side deployment and classroom interaction sequence; (b) preparation of the partially randomized pre-trained model in the Python environment; (c) smartphone-mediated data capture/labeling setup used in the activity.

3. Evaluation Method

3.1. Design and participants

We used a one-group pre-test/post-test quasi-experimental design to evaluate the educational effectiveness of the PWA workshop in a regular university class. The study was conducted in a first-year course for non-CS students in an interdisciplinary health and social well-being program. The intervention was a single in-class session in October 2025 after lecture content introducing basic AI/ML concepts.

After removing Qualtrics metadata rows and duplicate responses by IP address (keeping the latest response per IP), the dataset contained 43 pre-workshop and 33 post-workshop records. Anonymous pre/post responses were matched by IP address, yielding a paired analytic sample of 28 participants. In the matched sample, prior coding experience was limited (No (n=21); Yes (n=7)), and baseline self-rated AI understanding was mostly low (2 - Very little understanding (n=13); 3 - Some understanding (n=9); 1 - No understanding at all (n=4); 4 - Good understanding (n=2)). Institutional fast-track Research Ethics Committee (REC) approval was obtained (HKBU REC/25-26/0027).

3.2. Measures, procedure, and analysis

Data collection included: (1) a pre-workshop background survey, (2) a five-item conceptual quiz administered pre/post, (3) a 10-item post-workshop Likert questionnaire on ease of use, usefulness, and self-efficacy/attitude, adapted from technology acceptance constructs (Davis, 1989; Venkatesh et al., 2003), and (4) three open-ended feedback questions.

The five quiz items covered: the purpose of labeling data (Q1), the advantage of using a pre-trained model (Q2), interpretation of approximately 50% test accuracy in a binary classifier (Q3), the role of a testing set (Q4), and the effect of poor-quality training images on model performance (Q5).

The classroom procedure comprised pre-test, guided PWA activity, post-test, and short reflection. Quiz knowledge was not taught as explicit answers; students were encouraged to infer concepts through guided interaction. Analysis used paired-samples t-tests for matched total quiz scores, Wilcoxon signed-rank tests as a non-parametric check, descriptive statistics plus Cronbach's alpha for post-workshop subscales (using all deduplicated post responses), and a light thematic coding pass for open-ended feedback (Braun & Clarke, 2006).

4. Results

4.1. Conceptual learning gains

Conceptual quiz performance improved after the workshop. In the matched sample (n=28), the mean score increased from 2.89 (SD=1.47) pre-workshop to 3.71 (SD=1.38) post-workshop. The improvement was statistically significant, $t(27)=2.21$, $p=0.035$. The mean gain was 0.82 points (95% CI [0.06, 1.58]).

A non-parametric sensitivity analysis showed a consistent result (Wilcoxon signed-rank statistic=54.0, $p=0.030$). Item-level accuracy gains were strongest for the question about interpreting approximately 50% test accuracy in a binary classifier and for the question about the purpose of labeling data, which aligns with the workshop's emphasis on supervised learning and testing.

As shown in Figure 2(a), the total quiz score distributions shift upward after the workshop when visualized using pre/post violin-plus-box summaries with overlaid participant scores and mean ± 95% confidence intervals. Figure 2(b) complements this by showing item-level accuracy changes, highlighting the largest gains in labeling and interpretation of approximately 50% accuracy.

4.2. Perceived usability, usefulness, and attitude

Using all post-workshop survey responses after deduplication (n=33), students reported positive perceptions across all three constructs. Mean ratings were high for perceived ease of use (M=4.16, SD=0.80), perceived usefulness (M=4.19, SD=0.75), and self-efficacy/attitude (M=4.12, SD=0.76) on a 1-5 scale. These values suggest that the mobile interface was usable for novices and that students perceived clear learning value in the hands-on activity.

The three subscales also showed high internal consistency in the post-workshop sample (alpha=0.925 for perceived ease of use, alpha=0.926 for perceived usefulness, and alpha=0.913 for self-efficacy/attitude). Figure 2(c) shows the subscale score distributions and summary statistics. Although the sample is modest and the instrument was adapted for a specific classroom context, these results support the reliability of the post-workshop perception measures used in this study.

4.3. Qualitative feedback themes

Open-ended responses were generally positive and short. A dominant positive theme was engagement with the hands-on process of training and testing a model on a phone (e.g., comments highlighting that the activity was fun, easy to understand, and helpful for understanding AI). A second theme was the novelty of directly controlling data collection and observing model performance.

The most common difficulties involved capturing clear photos, deciding how to label data correctly, and understanding abstract concepts such as layers in machine learning and the relationship between data and model accuracy.

Improvement suggestions focused on clearer explanations of labeling and datasets, clearer label-to-class mapping in the app interface, more examples, and support for selecting images from the photo album instead of real-time capture only.

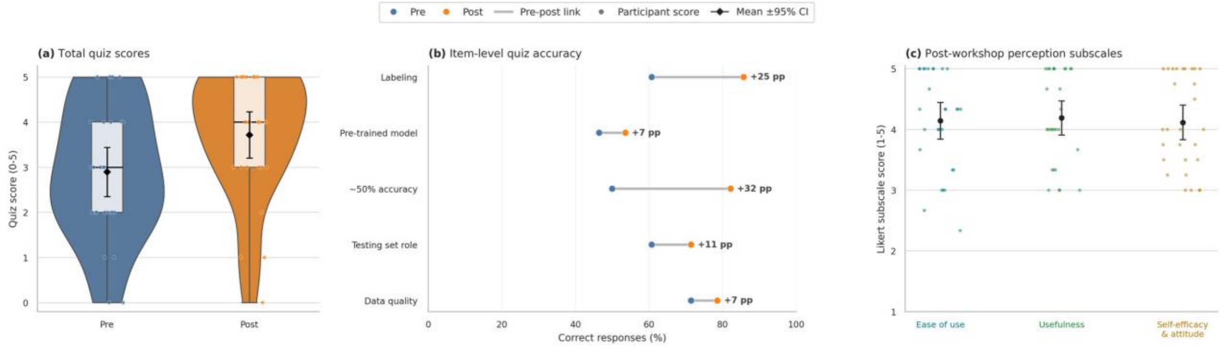


Figure 2. Quantitative outcomes of the workshop. (a) Distribution of matched pre/post total quiz scores with participant-level scores and mean ± 95% CI; (b) item-level pre/post accuracy and percentage-point gains; (c) post-workshop perception subscale score distributions and summary statistics.

5. Discussion

The results support the practical value of a code-free, bring-your-own-device approach for introductory AI education in interdisciplinary settings. The PWA reduced setup overhead while still allowing students to experience core steps of supervised learning, including collecting examples, assigning labels, training a model, and evaluating performance on withheld data. This is especially useful in classrooms where device heterogeneity, limited lab access, and varying technical backgrounds make conventional coding labs difficult to deploy.

An additional contribution is methodological, technical, and pedagogical: the study shows that a relatively lightweight PWA engineering approach can produce an AI learning tool without institution-managed compute infrastructure while supporting guided conceptual discovery. This framing aligns the workshop with guided-discovery and experiential learning approaches, because students do not receive quiz answers directly but instead build understanding by manipulating data, observing model behavior, and reflecting on the consequences for training and testing. The observed gains therefore suggest that the PWA can serve as an effective instructional scaffold for introductory ML learning, not merely as a deployment mechanism.

The strongest item-level improvements are pedagogically meaningful: students improved most on interpreting a near-random accuracy score and on the role of labeled data, both of which are difficult concepts for novices to internalize from definitions alone. At the same time, the evidence should be interpreted cautiously because the study used a one-group pre-test/post-test design, IP-based matching for paired analysis, and a single course context. Future work should compare this heuristic PWA workshop against lecture-only or computer-lab alternatives, test retention, and refine the interface (e.g., album upload and clearer label indicators) while preserving the low-cost deployment advantages of the PWA approach.

6. Conclusion

We presented a smartphone-enabled PWA and workshop design for teaching introductory ML concepts without coding and with minimal infrastructure requirements. Beyond the classroom intervention itself, the study contributes an implementation pattern showing how a comparatively simple browser-native application can operationalize transfer learning concepts for novice learners in a form that is practical for educators to deploy and adapt.

The evaluation with first-year non-CS students showed significant conceptual learning gains and strong post-workshop perceptions of usability and usefulness. Together, these findings support the role of lightweight PWAs as a credible technical foundation for widening participation in AI education across interdisciplinary and resource-constrained settings.

Acknowledgments

The authors thank the HSWB1005 students who participated in the classroom activity and survey. The study was conducted under HKBU REC/Ethics/Human (Non-clinical) Fast Track project REC/25-26/0027 (planned period 22 Sep 2025 to 21 Nov 2025; intended recruitment n=50; voluntary and anonymous participation with no incentives).

References

AI Index Steering Committee. (2022). AI Index Report 2022. Stanford Institute for Human-Centered Artificial Intelligence. https://hai.stanford.edu/assets/files/2022-ai-index-report_master.pdf

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, 13(3), 319–340. https://doi.org/10.2307/249008

Kermany, D. S., Goldbaum, M., Cai, W., Valentim, C. C. S., Liang, H., Baxter, S. L., McKeown, A., Yang, G., Wu, X., Yan, F., Dong, J., Prasadha, M. K., Pei, J., Ting, M. Y. L., Zhu, J., Li, C., Hewett, S., Dong, J., Ziyar, I., … Zhang, K. (2018). Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. Cell, 172(5), 1122-1131.e9. https://doi.org/10.1016/j.cell.2018.02.010

Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering, 22(10), 1345–1359. https://doi.org/10.1109/TKDE.2009.191

Smilkov, D., Thorat, N., Assogba, Y., Yuan, A., Kreeger, N., Yu, P., Zhang, K., Cai, S., Nielsen, E., Soergel, D., Bileschi, S., Corrado, G., Wattenberg, M., & Viégas, F. B. (2019). TensorFlow.js: Machine learning for the web and beyond. arXiv Preprint arXiv:1901.05350. https://arxiv.org/abs/1901.05350

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. MIS Quarterly, 27(3), 425–478. https://doi.org/10.2307/30036540

Hsu, T. C., & Ogata, H. (Eds.). (2026). *Proceedings of the 10th International Conference on Artificial Intelligence, Computational Thinking and STEM Education* .Kumamoto: The Asia-Pacific Society for Computers in Education.

Social and Humanoid Robots in the New Era of Educational Robotics

Antun DROBNJAK[1], Tomislav RAJNOVIC[2], Tomislav JAGUST[3] & Ivica BOTICKI[4*]
[1,2,3,4] Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia
*ivica.boticki@fer.hr

Abstract: *This paper presents results of a larger systematic review on the development of educational robotics from 2016-2025, with particular emphasis on its application in combination with virtual tools within the educational metaverse. Various studies are analyzed to identify the types of robots and virtual digital tools, while the collected data enable the identification of patterns in their application. The study places a special emphasis on the evolution of educational robotics in terms of the tools and approaches used by examining two adjacent periods: 2016-2020 and 2021-2025. The findings show a clear shift in educational robotics between those two periods, with growing emphasis on social robots, humanoids, simulation environments, and AI-driven tools. Virtual platforms and learning-based approaches have become central to development and experimentation, while hardware-intensive applications continue to face practical and scalability constraints. Overall, the field is increasingly shaped by the integration of AI and virtual technologies, with future progress depending on how effectively digital advances can be translated into reliable physical systems in educational settings.*

Keywords: educational robotics, metaverse, educational activities, review

1. Introduction

Robotics has rapidly expanded in research and consumer domains, with education as a key application. It is used from kindergarten to higher education, supporting STEM and non-STEM learning while fostering cognitive, collaborative, and problem-solving skills. Its use can be framed in three ways: as a learning objective, aid, or tool (Eguchi, 2021). As a multidisciplinary field combining robotics, pedagogy, and psychology, educational robotics seeks to enhance learning through pedagogically informed design (Angel-Fernandez & Vincze, 2018). While robots can increase curiosity and engagement, challenges such as cognitive load and skill complexity remain (Sapounidis & Alimisis, 2021).

At the same time, the educational metaverse has emerged as an immersive environment integrating augmented reality, virtual worlds, and related technologies. It blends physical and digital spaces, enabling real-time interaction independent of time and location (Zhang et al., 2022). Research shows that immersive technologies, especially when combined with learning analytics, can improve collaboration, motivation, and achievement (Lampropoulos & Evangelidis, 2025), though adoption depends on factors such as self-efficacy, effort expectancy, and immersion (Al-Kfairy et al., 2024). Although educational robotics, seamless learning, and the metaverse are widely reviewed, they are rarely examined together. While education is a leading domain of metaverse development (Herath et al., 2024), subject-specific and context-sensitive analyses remain limited. This fragmentation highlights a gap in understanding how to systematically integrate robotic and immersive technologies for seamless educational experiences.

This study examines the intersection of educational robotics and metaverse technologies, mapping their development from 2016 to 2025. It analyzes hardware and software trends across two periods through a systematic review of robotic platforms and virtual tools, identifying shifts in research focus and methods, and highlighting directions for integrating physical robots with virtual environments in education.

2. Methods

This paper employs a phased study selection Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) approach that consists of identification, screening, eligibility check, and inclusion phases to ensure a systematic and transparent approach in choosing the relevant studies for review (Liberati et al., 2009). The specific queries (see Table 1) were executed in October 2025 across the four major academic databases: the Web of Science (WoS) database, Scopus database, the Institute of Electrical and Electronics Engineers (IEEE) Xplore Digital Library and the Association for Computing Machinery (ACM) database. The initial search identified a total of 3,867 records. After removing 83 duplicate entries, 3,784 records remained for screening.

Table 1. *Search query*

WoS Query Example	TAK= ("educat* robot*") and ("metaverse" or "virtual reality" or "augmented reality" or "blockchain" or "artificial intelligence" or "extended reality" or "digital twin" or "machine learning" or "digital twins" or "mixed reality" or "avatar" or "deep learning" or "virtual worlds" or "internet of things" or "nft" or "vr" or "edge computing" or "virtual world" or "avatars" or "virtual reality (vr)" or "blockchains" or "web 3.0" or "iot" or "second life" or "6g" or "ai" or "augmented reality (ar)" or "industry 4.0" or "games" or "internet" or "cryptocurrency" or "three-dimensional displays" or "immersive technologies" or "reinforcement learning" or "non-fungible tokens" or "ar" or "e-learning" or "extended reality (xr)" or "cloud computing" or "social media" or "big data" or "real-time systems" or "xr" or "computer vision" or "smart contracts" or "artificial intelligence (ai)" or "6g mobile communication" or "nfts" or "chatgpt" or "generative ai" or "sensors" or "wireless communication" or "web3" or "gaming" or "sentiment analysis" or "simulation" or "smart cities" or "cybersecurity")

During the screening phase clearly defined inclusion and exclusion criteria were established to systematically determine which studies would be retained. These criteria were formulated to ensure that selected publications were directly relevant to the intersection of educational robotics and metaverse technologies, with particular emphasis on issues related to technological design and development. Two independent coders applied inclusion and exclusion criteria to a random sample of 300 papers based on titles and abstracts. Inter-coder agreement was substantial (Cohen's $\kappa = 0.650$). Discrepancies were resolved through discussion, after which both coders independently screened the remaining studies using the agreed criteria.

The subsequent phase of analysis followed a qualitative approach. Open coding was applied as the initial level of analysis to identify discrete concepts, categories, and emerging themes within the data. This was followed by axial coding, which served as a second, more integrative stage of analysis. Through axial coding, relationships among the previously identified codes were examined and organized, enabling refinement and consolidation of categories. As described by Strauss (1987), axial coding focuses on linking open codes to form central (core) categories, which emerge from clusters of strongly related or overlapping codes supported by substantial evidence.

In the subsequent coding phase, three coders jointly analyzed 5% of the papers to refine the coding scheme, reaching agreement before proceeding independently to ensure consistency and reliability. Following the open coding process, two main analytical dimensions emerged: Robot Types (e.g., humanoid, social, wheeled) and Virtual and Digital Technology Types (e.g., simulation systems, immersive technologies, AI based tools). The resulting structure forms the basis for a comparative analysis of two distinct time periods: 2016–2020 and 2021–2025. The division between these periods is analytically relevant because the early period reflects foundational integration of robotics with digital and immersive technologies, whereas the later period captures a phase marked by rapid advancements in artificial intelligence, increased accessibility of AR/VR systems, and accelerated digital transformation, particularly following the global shift toward remote and technology-mediated solutions after 2020. By comparing these intervals, the study aims to determine whether the field has introduced new robot categories, adopted more advanced intelligent systems, or shifted its technological focus. The final corpus of the review consisted of 721 included studies, following the screening and application of inclusion and exclusion criteria across the merged dataset from multiple academic databases.

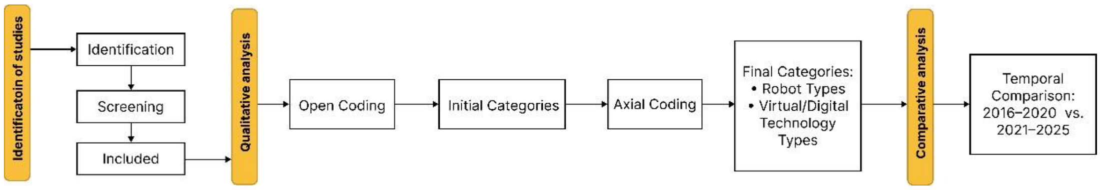


Figure 1. PRISMA-Informed Flow Diagram of Study Selection and Qualitative Analysis

3. Results

This section reports the main findings of the review, highlighting both the quantitative development of the field and shifts in its research composition. It first outlines overall publication trends and then examines changes in thematic and methodological focus.

3.1. Major Quantitative Shift (2016–2020 vs. 2021–2025

Clear structural shifts are observable across both physical robotic platforms and virtual technologies (see Figure 2). A more detailed explanation of these categories is provided in previous work (Boticki et al., under review). Among physical robot categories, service robots exhibit the most substantial growth, followed by social robots. Robotic arms also show strong expansion, while humanoid robots demonstrate a clear upward trend. In contrast, wheeled robots, although remaining the most prevalent platform across both periods, display comparatively moderate growth. These trends indicate that, while mobility-centered systems continue to serve as a foundational platform, recent research increasingly prioritizes socially interactive and service-oriented robotic applications. On the virtual technology side, the acceleration is even more pronounced. Algorithm-focused research shows the most dramatic expansion across all categories. Learning platforms and educational games also expand considerably, while augmented and virtual reality technologies demonstrate steady growth. Notably, AI tools and chatbots emerge as a distinct category only in the later period, reflecting the rapid integration of AI into robotics-related research. Simulation technologies show the largest overall expansion among virtual systems, underscoring a marked shift toward virtual experimentation and model-based validation.

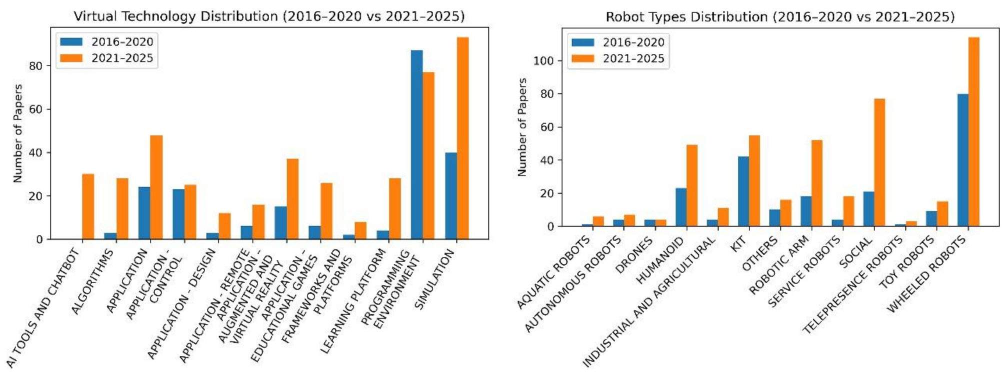


Figure 2. Distribution of Virtual Technologies and Robot Types

3.2. Shifts in Research Composition

The distribution of research themes changed considerably between the periods of 2016–2020 and 2021–2025 (show in Figure 2 and Figure 3). During 2016–2020, the literature was primarily dominated by wheeled robots and manipulation-focused systems (e.g., robotic arms) (e.g., (Escobar et al., 2020); (Farias et al., 2019)). In contrast, the 2021–2025 period reflects a more diversified structure, with social robots (e.g., (Al Hakim et al., 2024)) and humanoid platforms (e.g., (Oralbayeva et al., 2025)) occupying a substantially larger share of total publications. This shift suggests a growing emphasis on human–robot interaction, assistive applications, and socially embedded robotics. On the virtual axis, simulation (e.g., (Saivichit & Aphiratsakun, 2024)) and algorithmic research (e.g., (Moezzi et al., 2023)) now constitute a larger proportion of literature compared with the earlier period. The increasing share of AI-driven tools (e.g., (Pande & Mishra, 2025)), learning platforms (e.g., (Zhang et al., 2023)), and immersive technologies further indicates a methodological transition toward data-driven, intelligent, and hybrid physical–virtual research paradigms.

Decomposition of total growth (show in Figure 3) indicates that a relatively small number of categories account for the majority of expansion observed between the two periods. Specifically, social robotics, simulation technologies, algorithmic research, and learning platforms collectively contribute most of the overall increase in publications.

Although service robots and humanoid robots represent a smaller share of total growth in absolute terms, their strong upward trajectories suggest that these areas constitute rapidly emerging and strategically expanding research domains.

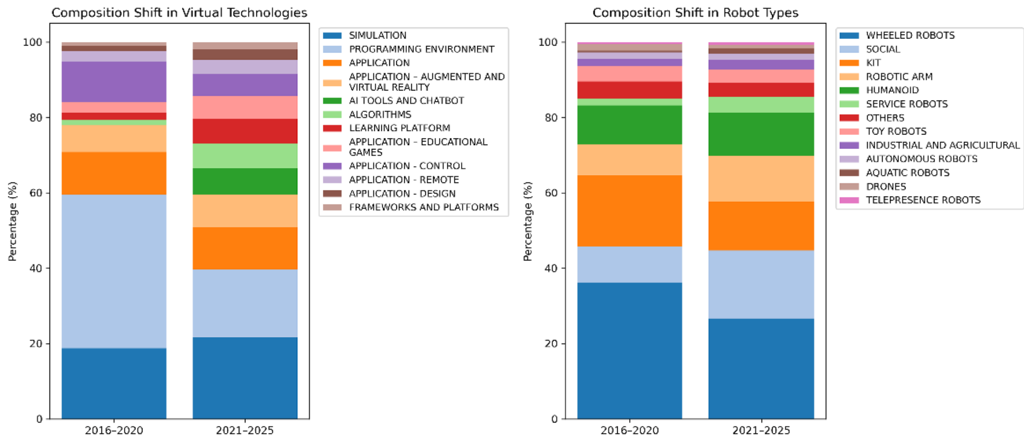


Figure 3. Composition shift in Virtual Technologies and Robot Types

4. Discussion

This section critically interprets the results by examining major shifts in robotics research and their broader significance. It explores the move toward interaction-centered systems, the uneven development of research domains, emerging trajectories, and the implications these trends hold for robotics education.

4.1. Towards Social and Humanoid Robotics

The results show not only growth, but a clear reorientation of the field. Between 2021 and 2025, robotics research shifted in how it defines its main problems, tools, and applications. Earlier work was largely centered on mobility and manipulation (Escobar et al., 2020; Farias et al., 2019). These areas remain important, but recent growth in social and humanoid robotics indicates a move toward interaction-centered systems (Al Hakim et al., 2024; Oralbayeva et al., 2025). Robots are increasingly assessed not only by mechanical performance, but by their ability to communicate, adapt, and operate in human environments. This reflects both technical progress and changing application needs.

Advances in technology and language processing have made socially responsive behavior more feasible, enabling robots to meet growing needs in education, therapy, and assistive settings.

At the same time, some technically complex domains, particularly advanced embodied manipulation, did not grow as strongly as might be expected. This likely reflects practical limits. Physical systems are costly, slower to iterate, and harder to scale than software-based research (Ligot & Birattari, 2022). Work in embodied intelligence requires extensive hardware testing and real-world validation. Part of the observed shift may therefore be driven not only by research interest, but by feasibility. The clearest trend is the growing centrality of AI-driven methods: algorithm-focused research, learning systems, and

AI tools have increased markedly and now occupy a larger share of literature. Rather than a separate topic, AI is increasingly embedded across robotics, reflecting its shift from an isolated field to a foundational component of technological development. In robotics, this appears in the rising reliance on data-driven and computational approaches in both development and experimentation. The rise of simulation further supports this shift.

Simulation is no longer just a testing tool; it has become a primary research environment (Saivichit & Aphiratsakun, 2024). Models are trained, refined, and evaluated in virtual settings before physical deployment. This reduces cost, speeds up iteration, and enables large-scale data generation (Ligot & Birattari, 2022). The research workflow itself has changed, with digital experimentation now preceding embodiment. Service robotics shows strong growth, likely driven by needs during and after the COVID-19 pandemic, with increased focus on healthcare, telepresence, and logistics. However, it does not dominate overall. This suggests that while external events can accelerate attention, sustained growth depends on long-term technical and infrastructural capacity (Vishwakarma et al., 2024).

4.2. Interpreting Uneven Growth Among the Identified Categories

An important finding is not only which categories expanded, but which did not. Growth is uneven across domains. One explanation lies in complexity. Research on full embodied autonomy, dexterous manipulation, and long-term deployment requires expensive hardware, extended testing, and interdisciplinary collaboration (Ligot & Birattari, 2022). These factors limit rapid output, while algorithmic research benefits from shared datasets, benchmarks, and scalable computing, enabling faster results and publication. Social and service robotics often require time-intensive, real-world studies, slowing visible growth. Overall, scalability drives this pattern: software-based research expands quickly, whereas hardware-intensive domains grow more slowly, reflecting structural conditions as much as conceptual priorities.

4.3. Emerging Research Trajectories and Educational Implications

The convergence of social robotics, AI integration, and simulation highlights key future directions. A central challenge is linking simulation-trained systems to reliable real-world interaction, making transfer from virtual to physical environments critical. Another priority is developing adaptable learning systems that move beyond task-specific solutions toward continual and transferable learning for humanoid and service robots. As robots become more socially embedded, evaluation must extend beyond technical performance to include usability, safety, trust, and long-term impact. Responsible deployment in healthcare and service contexts also requires greater attention to safety standards, ethics, and regulation.

These trends directly affect robotics education. The field is shifting from hardware-focused development to scalable, simulation- and data-driven approaches. While mobility and manipulation remain important, the growth of social and humanoid robotics emphasizes interaction and human-centered use. This creates a gap, as education often remains constrained by limited hardware and narrow curricula. To keep pace, robotics education must prioritize simulation-based learning, integrate physical–digital environments, and include AI and human–robot interaction competencies.

5. Conclusions

Robotics research (2016–2025) has shifted from mechanical systems to AI-driven, learning-based, and socially oriented applications, changing not only what robots are but how they are trained and evaluated through computational tools. Growth is strongest in scalable areas like simulation and data-driven methods, which now dominate experimentation, while hardware domains expand more slowly due to resource demands, longer testing, and real-world validation. A key challenge is transferring advances from virtual environments to reliable physical systems, especially in complex human settings. These trends highlight the need for robotics education to incorporate simulation-based environments, interdisciplinary content, and hybrid physical–digital approaches to support practical application.

References

Al Hakim, V. G., Yang, S.-H., Wang, J.-H., Lin, H.-H., & Chen, G.-D. (2024). Digital twins of pet robots to prolong interdependent relationships and effects on student learning performance. IEEE Transactions on Learning Technologies, 17, 1843–1857.

Al-Kfairy, M., Ahmed, S., & Khalil, A. (2024). Factors impacting users' willingness to adopt and utilize the metaverse in education: A systematic review. Computers in Human Behavior Reports, 15, 100459.

Angel-Fernandez, J. M., & Vincze, M. (2018). Towards a definition of educational robotics. Austrian Robotics Workshop 2018, 37. https://library.oapen.org/bitstream/handle/20.500.12657/39628/9783903187221.pdf#page=39

Eguchi, A. (2021). AI-Robotics and AI Literacy. In M. Malvezzi, D. Alimisis, & M. Moro (Eds.), Education in & with Robotics to Foster 21st-Century Skills (Vol. 982, pp. 75–85). Springer International Publishing. https://doi.org/10.1007/978-3-030-77022-8_7

Escobar, L., Moyano, C., Aguirre, G., Guerra, G., Allauca, L., & Loza, D. (2020). Multi-robot platform with features of Cyber-physical systems for education applications. 2020 IEEE ANDESCON, 1-6. https://ieeexplore.ieee.org/abstract/document/9272030/

Farias, G., Fabregas, E., Peralta, E., Vargas, H., Dormido-Canto, S., & Dormido, S. (2019). Development of an easy-to-use multi-agent platform for teaching mobile robotics. IEEE Access, 7, 55885–55897.

Herath, H. M. K. K. M. B., Mittal, M., & Kataria, A. (2024). Navigating the metaverse: A technical review of emerging virtual worlds. WIREs Data Mining and Knowledge Discovery, 14(4), e1538. https://doi.org/10.1002/widm.1538

Lampropoulos, G., & Evangelidis, G. (2025). Learning analytics and educational data mining in augmented reality, virtual reality, and the metaverse: A systematic literature review, content analysis, and bibliometric analysis. Applied Sciences, 15(2), 971.

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: Explanation and elaboration. Bmj, 339. https://www.bmj.com/content/339/bmj.b2700.abstract

Ligot, A., & Birattari, M. (2022). On Using Simulation to Predict the Performance of Robot Swarms. *Scientific Data*, *9*(1), 788. https://doi.org/10.1038/s41597-022-01895-1

Moezzi, R., Saw, A., Bischoff, S., Cyrus, J., & Hlava, J. (2023). Object Detection by Tiny-YOLO on TurtleBot3 as an Educational Robot. In H. Sharma, V. Shrivastava, K. K. Bharti, & L. Wang (Eds.), Communication and Intelligent Systems (Vol. 689, pp. 619–630). Springer Nature Singapore. https://doi.org/10.1007/978-981-99-2322-9_47

Oralbayeva, N., Telisheva, Z., Amir, A., Zhanatkyzy, A., Aimysheva, A., & Sandygulova, A. (2025). Moveable Älıpbi: Design of Montessori-Based Child-Robot Interaction for Long-Term Alphabet Learning. *International Journal of Social Robotics*, *17*(10), 2107–2122. https://doi.org/10.1007/s12369-024-01189-z

Pande, A., & Mishra, D. (2025). Humanoid robot as an educational assistant – insights of speech recognition for online and offline mode of teaching. *Behaviour & Information Technology*, *44*(5), 975–992. https://doi.org/10.1080/0144929X.2024.2344726

Saivichit, K., & Aphiratsakun, N. (2024). Robotic Kinematics Learning Guidance Based on Game Engine Simulation. 2024 9th International STEM Education Conference (iSTEM-Ed), 1–4. https://ieeexplore.ieee.org/abstract/document/10663147/

Sapounidis, T., & Alimisis, D. (2021). Educational Robotics Curricula: Current Trends and Shortcomings. In M. Malvezzi, D. Alimisis, & M. Moro (Eds.), Education in & with Robotics to Foster 21st-Century Skills (Vol. 982, pp. 127–138). *Springer International Publishing*. https://doi.org/10.1007/978-3-030-77022-8_12

Strauss, A. L. (1987). Qualitative analysis for social scientists. Cambridge university press. https://books.google.com/books?hl=hr&lr=&id=y16ww5ZsJ0AC&oi=fnd&pg=PR11&dq=Straus

Vishwakarma, L. P., Singh, R. K., Mishra, R., Demirkol, D., & Daim, T. (2024). The adoption of social robots in service operations: A comprehensive review. *Technology in Society*, *76*, 102441.

Zhang, X., Breazeal, C., & Park, H. W. (2023). A Social Robot Reading Partner for Explorative Guidance. Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction, 341–349. https://doi.org/10.1145/3568162.3576968

Zhang, X., Chen, Y., Hu, L., & Wang, Y. (2022). The metaverse in education: Definition, framework, features, potential applications, challenges, and future research topics. *Frontiers in Psychology*, *13*, 1016300.

Hsu, T. C., & Ogata, H. (Eds.). (2026). *Proceedings of the 10th International Conference on Artificial Intelligence, Computational Thinking and STEM Education* .Kumamoto: The Asia-Pacific Society for Computers in Education.

Reconstructing Computational Thinking in the AI Era: A Human-Centered Approach

Jin-Bao ZHANG
Faculty of Education, Beijing Normal University
Chinazhangjb@bnu.edu.cn

Abstract: *Artificial intelligence (AI) disrupts classical computational thinking (CT) by rendering its "instrumental rationality" theoretically bankrupt. Avoiding the "de-computation" risk of AI literacy alternatives, this paper proposes a Human-Centered Three-Dimensional CT Framework (HC-3D-CT). To manage the AI era's "triple cognitive load," HC-3D-CT integrates behavioral inclination (CT-BI) as its core, critical practice (CT-S) as the mediator, and value-laden knowledge (CT-K) as the foundation. This paradigm drives a pedagogical shift from skills-training to value-driven teaching, concluding with an empirical roadmap for measuring CT-BI to cultivate ethical innovators..*

Keywords: Computational Thinking; Human-Centered Framework; Responsibility Ethics; Value-Driven Pedagogy

1. Introduction

With the rapid advancement of generative AI, Computational Thinking (CT) faces a severe paradigm crisis. AI is fundamentally automating the classic CT foundation of "humans as designers" (Weintrop et al., 2016); Large Language Models now autonomously generate and optimize complex code, rendering traditional procedural programming tasks obsolete (Becker et al., 2023). Moreover, AI's "black box" nature defies classical decomposition skills.

This automation exposes the "theoretical bankruptcy" of traditional CT's underlying "instrumental rationality." By overemphasizing technical efficiency and marginalizing ethics, classical CT cannot address core AI challenges like algorithmic bias. Current academic responses—either a "patchwork" integration of AI or a complete replacement with "AI literacy"—fail to resolve this. The "patchwork" method retains the old paradigm's structural flaws, while the "replacement" theory risks "de-computationalization," depriving students of the cognitive foundation required to critically deconstruct AI systems.

To address these challenges, CT education must shift from procedural coding to the ethical evaluation and responsible management of AI. This paper reconstructs the CT paradigm from instrumental rationality to human-centered thinking. By establishing a "dual necessity" standard across philosophical and cognitive dimensions, we propose the Human-Centered Three-Dimensional Computational Thinking Framework (HC-3D-CT) to reshape value-driven pedagogy and cultivate future ethical innovators.

2. The "Dual Necessity" for Theoretical Reconstruction

Before constructing a new framework, we must answer: What "non-negotiable" theoretical criteria must a Computational Thinking (CT) framework in the AI era satisfy? This paper argues that there exists a "dual necessity": it is both an inevitable requirement of the philosophy of technology and a necessary choice of cognitive science.

2.1. The Necessity of Philosophy: From Instrumental Rationality to Responsibility Ethics

The ontological basis of classical CT theory of instrumental rationality is from seeing technology as a neutral, efficient tool. But AI's non - neutrality and its ability to reshape social structures can't fit this old paradigm. Heidegger

(1977) warned that viewing the world through a technological lens obscures its humanistic essence. Due to AI's irreversible and global consequences, Hans Jonas's (1984) "ethics of responsibility" provides the needed theoretical pivot: ethical responsibility should be the primary principle of technological practice, not just an afterthought.

Therefore, the central challenge of the AI era has shifted from technical execution (how to realize) to value-based problems (why to realize, and how to realize responsibly). From this, we draw the following conclusion: Criterion One of the new paradigm demands a human-centered turn in philosophy, positioning ethical responsibility and human-centered awareness (CT-BI) as its structural core, rather than as an external add-on.

2.2. The Necessity of Cognition: Managing "Triple Cognitive Load"

Learning in the AI era goes beyond simple procedural knowledge acquisition. It is a highly concurrent process involving cognition, strategy, and situational value judgments. The classical CT framework mainly focuses on procedural cognitive load (e.g., algorithms and programming languages), but the "black box" nature and ethical complexities of AI bring two heavier cognitive loads that are blind spots in traditional frameworks.

- Strategic Cognitive Load: In the face of AI uncertainty, students must shift from simply "knowing how to use" to "knowing how to evaluate," which requires high-level metacognitive abilities.
- Contextual Load: This is the most critical load in the AI era. Students must navigate ambiguous information, ethical dilemmas, and multi-objective trade-offs (e.g., efficiency vs. fairness) in real-world contexts.

Structurally, the traditional CT framework can't manage strategic and contextual loads. So, observed teaching phenomena like "ethical avoidance" are not individual student failures but inevitable cognitive symptoms of the old paradigm. Thus, we conclude that Criterion Two of the new paradigm is to provide a "human - centered cognitive ecology" to systematically manage the "triple cognitive load".

2.3. Systematic Comparison of Frameworks

To clarify HC-3D-CT framework's unique contribution, it's essential to systematically compare it with existing paradigms. Table 1 outlines the structural differences between the proposed framework and three mainstream approaches.

Table 1. *Systematic Comparison of CT and AI Education Frameworks*

Framework	Core Philosophy	Primary Focus	Structural Limitations	HC-3D-CT's Breakthrough
Traditional CT (e.g., Brennan & Resnick, 2012)	Instrumental Rationality	Coding & algorithmic efficiency	Overemphasizes skills; lacks systematic ethical structure.	Embeds ethics as a core structural element.
AI Literacy (e.g., Long & Magerko, 2020)	User-Centric Pragmatism	AI applications & social impacts	Risks "de-computationalization" from a user perspective.	Maintains CT cognitive foundation (abstraction, modeling).
Critical CT / HCC	Critical Pedagogy	Social justice & human-computer interaction	Lacks a cognitive model to integrate skills and ethics.	Lacks a cognitive model to integrate skills and ethics.
HC-3D-CT (Proposed)	Responsibility Ethics	Value-driven practice & ethical decision	N/A	

3. The Human-Centered 3D Framework: An Integrative Response

The Human-Centered Three-Dimensional Computational Thinking Framework (HC-3D-CT) is an integrative response to the "dual necessity" of the AI era. It shifts the paradigm from "how to compute efficiently" to "why and how to compute responsibly". Rather than a simple addition of dimensions, HC-3D-CT forms an "intrinsic—outreach" cognitive ecosystem designed to structurally embed responsibility ethics and systematically manage the threefold cognitive load.

3.1. CT-BI (Behavioral Inclination): The "Human-Centered Core" and Situational Load

Manager The CT-BI dimension is the theoretical core and teleological endpoint of the framework. Responding directly to the philosophical necessity, it anchors the framework's value orientation, structurally translating the "ethics of responsibility" (Jonas, 1984) into educational practice. Cognitively, CT-BI acts as the manager of "situational load," providing students with the cognitive framework and motivational source to handle real-world ethical dilemmas.

It must be emphasized that CT-BI is philosophically distinct from general non-cognitive skills (such as grit). While grit might manifest as working hard to debug code until it runs successfully, CT-BI is a tendency toward the "non-neutrality" of technology . For example, a student demonstrating CT-BI would not just ensure the code works; upon recognizing potential algorithmic bias (CT-K), they would actively question whether the program should be deployed at all . Its hierarchical indicators form a maturity model for ethical practice: from L1 "ethical identification," to L2 "weighing and assessment," and finally to L3 "critical reflection" (see Table 2).

Table 2. *Tiered Indicators of Ethical Responsibility (CT-BI)*

Level	Core Characteristics	Observable Assessment Evidence
L1: Ethical Identification	Identifies basic technology ethics issues (e.g., algorithmic bias).	Highlights potential bias factors (e.g., gender/race) in provided datasets during class discussions.
L2: Trade-off Evaluation	Balances technical efficiency with social fairness; articulates value conflicts.	Submits comparison logs detailing trade-offs between algorithmic efficiency and its impact on fairness.
L3: Critical Reflection	Reflects on structural social impacts; proactively embeds human-centered values into design.	Proactively introduces data diversity in projects and authors a brief "Technology Social Impact Assessment" report.

3.2. CT-S (Skills): The "Critical Practice" Pathway and Strategic Load Guide

The "dimensionality reduction strike" of AI against classic CT skills forces the CT-S dimension to shift from "technical operation" to "critical praxis". Cognitively, CT-S serves as the guide for "strategic load," directing students to shift cognitive resources from procedural coding toward evaluation and reflection.

Within this framework, the skill of "assessment and reflection" acquires a new philosophical status: it is the core methodology for practicing the ethics of responsibility (Jonas, 1984). The question is no longer "Can it run?" but "Should it run?". Similarly, "data practice and inquiry" becomes an approach to critically explore the human reality obscured by data representations. CT-S is thus transformed from an automatable metacognitive strategy into a value-laden "human-centered practice".

3.3. CT-K (Knowledge): The Cognitive Foundation of "Ethics of Responsibility"

The reconstruction of the CT-K dimension is a theoretical response to the AI "black box crisis". Cognitively, it serves as the foundation that supports the preceding two dimensions. The HC-3D-CT framework insists that CT-K itself must be "value-laden".

Therefore, incorporating "algorithmic bias," "data privacy," and "social impact assessment" into CT-K is not meant for technical debugging, but for philosophical "un-concealment". Its purpose is to reveal how the logic of technical frameworks systematically conceals social justice and human values. This knowledge system enables "responsibility ethics" to become possible in practice.

4. Pedagogical-Ontological Shift: Towards "Value-Driven" Teaching

The HC-3D-CT framework shifts pedagogy from technical execution to "value-driven" ethical exploration. This is illustrated when students design an AI traffic algorithm for a divided city. First, students activate CT-K by learning machine learning alongside "algorithmic bias," recognizing algorithms as value-laden. Next, applying CT-S, they move beyond assessing running speed to multi-dimensionally compare models prioritizing absolute efficiency versus fair waiting times for marginalized communities. Finally, achieving CT-BI, students confront the efficiency-fairness conflict, prudently sacrificing marginal efficiency for social equity and submitting logs to justify their ethical trade-offs. Thus, technical practice effectively overturns "instrumental rationality."

5. Conclusion and Future Roadmap

Responding to CT's paradigm crisis, this paper proposes the HC-3D-CT framework. Shifting from "instrumental rationality" to "responsibility ethics" (Jonas, 1984), it centers on ehavioral Inclination (CT-BI), integrating critical practice (CT-S) and value-laden knowledge (CT-K) to manage the AI era's triple cognitive load. To achieve empirical maturity, we propose a three-step roadmap: 1) develop scenario-based rubrics to capture ethical decision-making tendencies; 2) conduct cognitive experiments to validate internal processing mechanisms; and 3) test pedagogical efficacy via classroom interventions. Ultimately, this human-centered paradigm can be extended to disciplines like data science or bioengineering to broadly address the universal crisis of instrumental rationality.

References

Becker, B. A., et al. (2023). Programming is hard—or at least it used to be: Educational opportunities and challenges of AI code generation. In Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1 (SIGCSE 2023). https://doi.org/10.1145/3545945.3569759.

Brennan, K., & Resnick, M. (2012). New frameworks for studying and assessing the development of computational thinking. In Proceedings of the 2012 Annual Meeting of the American Educational Research Association (AERA). Vancouver, Canada.

Heidegger, M. (1977). The question concerning technology and other essays (W. Lovitt, Trans.). Harper & Row.

Jonas, H. (1984). The imperative of responsibility: In search of an ethics for the technological age. University of Chicago Press.

Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (pp. 1–16). Honolulu, HI. https://doi.org/10.1145/3313831.3376749.

Weintrop, D., Beheshti, E., Horn, M., Orton, K., Jona, K., & Wilensky, U. (2016). Defining computational thinking for mathematics and science classrooms. Journal of Science Education and Technology, 25(1), 127–147. https://doi.org/10.1007/s10956-015-9581-5..

Hsu, T. C., & Ogata, H. (Eds.). (2026). *Proceedings of the 10th International Conference on Artificial Intelligence, Computational Thinking and STEM Education* .Kumamoto: The Asia-Pacific Society for Computers in Education.

Exploration of Workforce Skills and Competency Requirements Among TVET Graduates

Muhd OMAR[1*], Yucan LI[2],Norhayati ABU HASSAN[3] & Rehan AHMAD[4]
[1,2,3,4] aFaculty of Educational Studies, Universiti Putra Malaysia, Malaysia
[1]Kuala Lumpur University of Science and Technology, Malaysia

Abstract: *Despite the increasing emphasis on Technical and Vocational Education and Training (TVET) in Malaysia, significant challenges persist in the job mismatch between the skills acquired by TVET graduates and the demands of the evolving job market. Many graduates face difficulties in securing employment due to a lack of alignment between their acquired skills and industry requirements, leading to underemployment or unemployment rates among TVET graduates. Surprisingly, data from a TVET graduate tracer study published by the Malaysian Ministry of Higher Education in 2020 indicates that 67.2 percent (6,243 graduates) were unemployed, with 7.5 percent uncertain about whether to wait for employment or pursue higher education. Personal issues, including family matters (6.5%; 600 graduates) and unclear career pathways (4.5%; 422 graduates), were among the unstated reasons why TVET graduates failed to enter the job market. On another note, 3.9% (365 graduates) reported a job mismatch of their qualifications, which rarely happens in TVET programs due to specialisations in skill-based industries. A qualitative within a phenomenological approach was employed to address the following objectives: (a) explore factors related to job mismatching among TVET graduates and (b) identify patterns of employment intake from employers specifically for TVET graduates. The informants for this study involved three Small Medium Enterprises (SMEs) and two Multi-national Corporations (MNCs) human resource manager in Malaysia. Research outcomes highlighting the barriers in improving employability among diverse backgrounds of TVET graduates and inspire initiatives promoting inclusivity and equality in access to job opportunities.*

Keywords: Employability skills, TVET, job market

1. Introduction

Thousands of graduates emerge from public and private higher education institutions. However, despite this accomplishment, many struggle to secure employment, often opting for different career paths or remaining unemployed while awaiting opportunities. The unemployment rate and employment difficulties among TVET graduates are increasingly concerning. In Malaysia, the disquieting challenge of fresh TVET graduate unemployment is highlighted by a staggering 67.2% unemployment rate, reported by the Ministry of Higher Education (2020) through a graduate tracer study. The trend of escalating unemployment among graduates across various programs has persisted. The Department of Statistics reports that this year, over 573.1 thousand graduates are part of the unemployed labor force, although the disparity has steadily decreased to 3.4% (Department of Statistics Malaysia, 2023).

Despite the increasing emphasis on Technical and Vocational Education and Training (TVET) programs worldwide, a concerning trend persists wherein a significant proportion of TVET graduates face challenges in securing suitable employment opportunities. Despite possessing specific skill sets and technical competencies gained from their educational training, a notable portion of these graduates encounter obstacles in entering the workforce effectively. This issue raises concerns about the misalignment between the skills acquired through TVET programs and the demands of the contemporary job market (Sibiya et al., 2020). The gap between the skills acquired during TVET education and the evolving needs of industries creates a scenario where graduates struggle to meet the job market's expectations, resulting in increased unemployment rates among this cohort (Singh & Tolessa, 2019). Additionally, factors such as outdated curricula, inadequate industry engagement, and insufficient soft skill development within TVET programs contribute significantly to the unemployability issue among TVET graduates (Yamada, & Otchia, 2021).

2. Issues Pertaining to TVET Employability

Limited collaboration between TVET institutions and industries exacerbates the divide between skills taught in classrooms and those required by employers. From the perspective of Human Capital Theory (Becker, 1964), such misalignment reduces the returns on educational investments because graduates are not equipped with competencies that enhance productivity and economic value in the labour market. Stronger partnerships between academia and industry are therefore imperative to ensure that TVET graduates possess relevant, up-to-date skills that match current market demands. Furthermore, there often exists a mismatch between the skills possessed by TVET and engineering education graduates and those demanded by industries. This mismatch frequently results in high rates of unemployment or underemployment among graduates, consequently hindering economic growth an outcome consistent with the human capital underutilisation described in Human Capital Theory.

The deficiency in soft skills and practical experience among graduates has been addressed in studies by Othman et al. (2018) and Tan et al. (2019). Employers often highlight the importance of communication skills, critical thinking, and adaptability, which are sometimes lacking in fresh graduates. In Social Cognitive Career Theory terms, the absence of these competencies can lower graduates’ confidence in their ability to perform in diverse job contexts, limiting both their application breadth and career advancement potential.

Embedding these soft skills within TVET curricula, alongside targeted career guidance, could therefore strengthen both the human capital value and self-efficacy of graduates, enhancing their long-term employability.

Thus, understanding the root causes of the unemployability problem among TVET graduates is crucial in devising effective interventions and strategies to bridge the gap between education and job marketability.

3. Method

As the research adopts a qualitative design, an exploratory approach grounded in induction becomes imperative. To meet this criterion, we gathered qualitative data directly from employers. This information sourced from the list of registered companies available in the Malaysia External Trade Development Corporation (MATRADE) database accessible at https://www.matrade.gov.my/en/directory-hub/malaysia-products-directory. Subsequently, Human Resource Managers were invited to participate in the interview sessions. Employing a phenomenological study enabled us to delve into employers’ perspectives during the hiring process of TVET graduates. The primary dataset was gleaned from these interview sessions, forming the basis to discern the factors influencing the recruitment of TVET graduates. Subsequently, thematic analysis was employed to identify generic competencies and domains that support job seekers in their applications.

Understanding the reticent nature of HR managers in certain industries, who tend to withhold data, is anticipated to pose a significant challenge in this study. Employing purposive sampling facilitated the acquisition of qualitative data from employers situated in Kuala Lumpur and Selangor serve as limitation of the study. A semi-structured interview was employed to elicit comprehensive narratives from these employers, encompassing industries beyond just services and manufacturing. Additionally, approval letters supporting and permitting the study's execution, provided by JKEUPM (Ethic Committee For Research Involving Human Subject) Reference No. JKEUPM-2024-842 was accompanied during the interview. Subsequently, narratives collected undergone analysis using the Atlas.Ti software.

4. Findings and Discussion

This section presents the results of the thematic analysis conducted from interview transcripts with industry employers involved in the recruitment and training of TVET graduates. Three Small Medium Enterprises (SMEs) and two Multi-national Corporations (MNCs) human resource managers were involved in estimated 1 hour to 1.5 hour of interview session.

Five key themes emerged: (1) Skills-Attitude Mismatch, (2) Curriculum-Industry Misalignment, (3) Limited Career Awareness, (4) Preference for Specialized Skills, and (5) Informal and Location-Based Hiring Practices. These five themes illustrate a multidimensional challenge in matching TVET graduates with employment. The findings point to a need for integrated efforts that bridge curriculum with practical realities, enhance student exposure to industrial pathways, and address the reliance on informal hiring processes. More importantly, employers' preference for attitude and specialisation over general qualifications signals a shift in the employability paradigm within technical industries.

Table 1. *Summary of the findings*

Theme	Description	Sample Quotes	Connection to RQs
1. Skills-Attitude Mismatch	TVET graduates often show weak discipline, poor attendance, and lack of perseverance despite having skills.	"You have skill but don't have attitude, it's difficult." "They only come "They only come	RQ1: Factors contributing to job mismatch
2. Curriculum-Industry Misalignment	A gap exists between what is taught in vocational colleges and what industries require in practice.	"They learn theory, but can't apply it." "They don't even know how to measure properly."	RQ1 & RQ2: Mismatch and employer intake decision factors
3. Limited Career Awareness	Graduates choose jobs based on proximity, not on skill relevance or career path, leading to mismatching.	"They choose internships close to home." "They don't know what companies like ours do."	RQ1: Job mismatch due to decision- making by graduates
4. Preference for Specialized Skills	Employers prefer modular, hands-on skills and are willing to pay more for certified, specialized workers.	"We want CNC skills, not just skills, not just certificates." "We'll pay more if they can do Aluminium welding."	RQ2: Patterns in hiring preferences by employers
5. Informal and Location-Based Hiring	Most SMEs rely on local contacts and informal platforms like Facebook or word of mouth for recruitment.	"We hire from around the area." "We post job offers in Facebook welder groups."	RQ2: Employment intake mechanisms among employers

This study explored the factors contributing to job mismatching among TVET graduates and the employment intake patterns among employers. The findings indicate that beyond technical competencies, employers value behavioral attributes, practical readiness, and proximity—all of which impact hiring decisions and graduate employability.

The mismatch between employers' expectations and graduates' workplace behavior indicates the importance of non-cognitive attributes in job readiness. Employers in this study repeatedly emphasized the primacy of discipline, work ethic, and perseverance over technical qualifications. This supports prior research which suggests that soft skills, particularly attitude and reliability, are pivotal in predicting job success in the vocational sector (Heinz, 2016 (di reference 2003); Lambert et al., 2020). (mungkin boleh tukar “Ismail et al., 2023). Although TVET programs emphasize technical mastery, these results call for greater attention to the development of affective attributes, such as accountability and resilience.

Another recurring issue was the disconnect between institutional training and industry application. Many employers reported that graduates lacked practical know-how even in core tasks such as machine setup and blueprint interpretation. This finding is consistent with past studies highlighting the inadequacy of skill transfer from training institutions to industrial contexts (Papier, 2017; Salleh et al., 2021). This gap requires businesses to bear the burden of retraining to bridge the shortfall in hands-on competence. Research highlights that while theoretical knowledge provides a crucial foundation, its translation into practical skills is essential for effective performance in real-world settings (Smith & Jones, 2022). For example, novice learners frequently find blueprint reading challenging due to the need to visualize complex structures from two-dimensional plans, interpret precise measurements, and apply relevant industry tasks that require more than conceptual understanding alone (Lee et al., 2021). To bridge this gap, stronger industry-academic partnerships are needed, with emphasis on co-developing modular curricula that reflect workplace realities (OECD, 2023).

Recent literature reinforces that many curricula still emphasize concepts over experiential learning, failing to equip students with the necessary hands-on skills (Brown & Green, 2020). Pedagogical studies emphasize that practical application through simulations, role-playing, and real-world problem solving strengthens skill retention and builds confidence (Davis, 2019). Hands-on experience transforms abstract knowledge into tangible competence, which is key to better outcomes in employment and productivity (Nguyen & Patel, 2023).

Graduate choices were often dictated by geographical convenience rather than long-term employability. This suggests a lack of structured career guidance, as also observed in other studies on TVET pathways in Southeast Asia (Martinez-Fernandez & Choi, 2012). The absence of institutional frameworks that support informed decision-making hinders students from accessing roles aligned with their interests and competencies. Career development initiatives should thus be integrated within the TVET ecosystem, particularly in the early stages of training (Cheong & Lee, 2016).

Employers expressed a preference for candidates with specialized, modular skillsets, such as CNC machining or aluminium welding. These preferences align with global trends that recognize micro-credentials and short-cycle certifications as highly employable formats, particularly in fast-changing labor markets (UNESCO-UNEVOC, 2020). The results highlight the need for flexible certification pathways that allow students to stack skills progressively and demonstrate real-world competencies (Wheelahan & Moodie, 2021). Hiring practices were shown to be highly localized and informal, with employers relying on word-of-mouth and social media platforms. While this approach is cost-effective, it reinforces asymmetrical access to employment, particularly disadvantaging students outside of geographic proximity. Prior research (Grollmann, 2018) has also noted the informal recruitment channels as a feature of SMEs, suggesting that institutional intermediaries (e.g., job-matching portals or career centers) may be needed to increase visibility and mobility for TVET graduates.

5. Conclusion

This study provides empirical insights into the complex challenges TVET graduates encounter as they transition into the workforce. While technical proficiency remains a critical foundation, it is increasingly evident that employers prioritize qualities such as work ethic, behavioural maturity, and an understanding of local hiring dynamics. The pervasive disconnect between institutional training and industry demands calls for a comprehensive revamp of TVET programs to move beyond theoretical knowledge to embed practical specialization, foster essential soft skills, and cultivate meaningful partnerships with employers. By embracing this multifaceted approach, TVET systems can better equip graduates to meet real-world expectations, thereby minimizing job mismatches and enhancing overall workforce readiness. Such evolution is not merely beneficial but necessary for bridging the gap between education and employment in today’s rapidly changing labour market.

Acknowledgements

This study is supported financially by the Fundamental Research Grant Scheme (FRGS) of the Malaysian Ministry of Education FRGS/1/2024/SSI09/UPM/02/5. The assistance of the Research Management Center (RMC) of Universiti Putra Malaysia in coordinating and distributing funds for this research is greatly appreciated.

References

Ali, M., Egbetokun, A., & Memon, M. H. (2018). Human capital, social capabilities and economic growth. Economies, 6(1), 1–18. https://doi.org/10.3390/economies6010002

Awad, A., & Youssof, I. (2016). The impact of economic globalisation on unemployment: The Malaysian experience. The Journal of International Trade & Economic

Development, 25(7), 938–958. https://doi.org/10.1080/09638199.2016.1151069 Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research

in Psychology, 3(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Chen, M., Sun, P., & Wan, Z. (2019). Matching supply and demand with mismatch-sensitive players. Available at SSRN 3458673.

Cheong, K. C., & Lee, H. A. (2016). Malaysia's Education Crisis – Can TVET Help?

Malaysian Journal of Economic Studies 53(1), 115-134. Retrieved from https://mjes.um.edu.my/article/download/2782/969

Department of Statistics Malaysia. (2023). Statistik guna tenaga. https://www.dosm.gov.my/portal-main/release-content/0803be88-8b7d-11ed-96a6-1866daa77ef9

Grollmann, P. (2018). The role of SMEs in vocational education and training in selected ASEAN member states. ILO Working Paper. https://www.ilo.org/

Heinz, W. R. (2003). From work trajectories to negotiated careers: The contingent work life course. In Handbook of the life course (pp. 185-204). Boston, MA: Springer US.

Ismail, A., Abdullah, R., & Rashid, R. A. (2023). Employers' perspectives on soft skills among Malaysian TVET graduates. Journal of Technical Education and Training, 15(1), 15–28. https://doi.org/10.30880/jtet.2023.15.01.002

Ministry of Higher Education (2020). Laporan Kajian Pengesanan Graduan TVET. https://great.mohe.gov.my/penerbitan/LAPORAN%20KAJIAN%20PENGESANAN%20 GRADUAN%20TVET%202020.pdf

Lambert, M., Vero, J., & Zimmermann, B. (2012). Vocational training and professional development: a capability perspective. International Journal of Training and Development, 16(3), 164–182. https://doi.org/10.1111/j.1468-2419.2012.00402.x

Lee, S., Kim, J., & Park, H. (2021). Blueprint reading difficulties among novice learners: A cognitive perspective. Construction Education Review, 7(4), 200-215.

Lim, S. M., Foo, Y. L., Yeo, M. F., Chan, C. Y. X., & Loh, H. T. (2020). Integrated work study program: Students' growth mindset and perception of change in work-related skills.

International Journal of Work-Integrated Learning, 21(2), 103–115.

Martinez-Fernandez, C., & Choi, K. (2012). Skills Development Pathways in Asia. OECD Local Economic and Employment Development (LEED) Papers, 2012(12). https://doi.org/10.1787/5k94hdlll7vk-en

OECD. (2023). Strengthening skills systems in Southeast Asia: Navigating policy alignment and industry engagement. OECD Publishing. https://doi.org/10.1787/af3e2836-en

Othman, Z., Shan, S. W., Yusoff, I., & Kee, C. P. (2018). Classification techniques for predicting graduate employability. International Journal on Advanced Science, Engineering and Information Technology, 8(4-2), 1712-1720.

Papier, J. (2009). Getting the Right Learners into the Right Programmes: An Investigation into Factors that Contributed to the Poor Performance of FET College Learners in NCV 2 and NCV 3 Programmes in 2007 and 2008–Reasons and Recommendations. Cape Town: FETI, University of the Western Cape.

Salleh, N. M., Emelifeonwu, J., Winterton, J., & Chan, K. M. (2017). Work-readiness in Malaysia. In Transitions from education to work (pp. 84-104). Routledge.

Sibiya, A. T., Nyembezi, N., & Bogopa, D. (2021). TVET engineering students' perceptions of the value of their qualification and the prospects of employment. Journal of Vocational, Adult and Continuing Education and Training, 4(1), 16. https://doi.org/10.14426/jovacet.v4i1.189

Siddiqa, A., Hussain, T., Qasim, M., & Javed, M. I. (2018). Siddiqa, A. ., Hussain, T. ., Qasim, M. ., & Javed, M. I. . (2018). The Impact of Globalization on Unemployment and Economic Growth: Panel Data Analysis for Developing Countries. Bulletin of Business and Economics (BBE), 7(3), 122-131. Retrieved from https://bbejournal.com/BBE/article/view/166

Singh, B., & Tolessa, M. B. (2019). TVET-industry linkage and collaboration in Ethiopia: A necessity for improving employability skill. International Research Journal of Engineering and Technology, 6(11), 3526–3532.

Smith, J., & Jones, L. (2022). Aligning educational outcomes with industry needs: A focus on practical skills. Vocational Training Quarterly, 21(1), 5-20.

Tan, L. M., & Laswad, F. (2019). Key employability skills required of tax accountants. Journal of the Australasian Tax Teachers Association, 14(1), 211–239.

UNESCO-UNEVOC. (2020). Trends shaping TVET: The future of technical and vocational education and training in a changing world. https://unevoc.unesco.org/

Wheelahan, L., Moldie, G., & Buchanan, J. (2012). Revitalising the 'vocational' in flows of learning and labour. Adelaide: NCVER.

Yamada, S., & Otchia, C. S. (2021). Perception gaps on employable skills between technical and vocational education and training (TVET) teachers and students: the case of the garment sector in Ethiopia. Higher Education, Skills and Work-Based Learning, 11(1), 199–213. https://doi.org/10.1108/HESWBL-08-2019-0105

Zainal, N., Law, S. H., Md Nassir, A., & Bakri, M. H. (2020). Does the Malaysian Islamic money market efficiently predict inflation in the future? Evidence from the Fisher effect theory. International Journal of Innovation, Creativity and Change, 11(3), 665-682. Retrieved from https://ijicc.net/images/vol11iss3/11332_Zainal_2020_E1_R.pdf

Hsu, T. C., & Ogata, H. (Eds.). (2026). *Proceedings of the 10th International Conference on Artificial Intelligence, Computational Thinking and STEM Education* .Kumamoto: The Asia-Pacific Society for Computers in Education.

Reinforcement Strategies in ChatGPT-Assisted Computational Thinking Education: A Comparison of Handwritten Practice and Oral Presentation

Chia-Hao CHIU*
Soochow University, Taipei, Taiwan
*starklab2020@scu.edu.tw

Abstract: *Generative AI tools such as ChatGPT are increasingly integrated into computational thinking and computer science (CS) education; however, concerns remain regarding the durability of learning outcomes. This study employed a quasi-experimental design with 120 undergraduates across two parallel ChatGPT-assisted CS1 sections differing only in reinforcement strategy: repetitive handwritten practice versus oral presentation. Results showed no significant group differences at baseline (M = 69.48 vs. 70.82, p = 0.78) or midterm (M = 76.7 vs. 74.4, p = 0.80). On the AI-free final exam, however, the handwritten-practice group scored significantly higher (M = 52.3 vs. 41.8; p = 0.01; Cohen's d = 0.70), with a more favorable score distribution (fewer below 50: 45.9% vs. 63.2%; more at 60+: 31.1% vs. 8.8%). These findings suggest that durable learning in AI-assisted environments depends more on instructional design than on AI access alone.*

Keywords: Generative AI; ChatGPT; Computational Thinking; CS Education; Repetitive Practice

1. Introduction

Since the public release of large language model (LLM) applications such as ChatGPT in late 2022, generative AI has rapidly influenced higher education, especially in STEM and computer science (CS) education. Prior studies show that AI tools can support computational thinking through instant explanations, debugging assistance, and adaptive feedback, often improving short-term performance and engagement in introductory computer science (CS1) contexts [1]–[3]. However, evidence on long-term learning remains mixed.

Although AI-assisted learning may produce immediate gains, assessments conducted without AI support often reveal limited retention and weaker independent problem-solving ability [4], [5]. These concerns suggest that effective AI integration depends not only on access to AI, but also on instructional design [7].

Two perspectives are especially relevant. Repetitive practice emphasizes active recall and procedural fluency [6], whereas learning-by-teaching highlights conceptual understanding through explanation [8]. Although both strategies are well established in STEM education, their interaction with generative AI remains underexplored. Few studies have directly compared these approaches within the same computational thinking context while holding AI access constant.

To address this gap, this study examined two parallel ChatGPT-assisted CS1 sections that were identical in course structure, materials, and assessments, differing only in reinforcement strategy:

- Handwritten repetitive practice (procedural fluency, active recall)
- Oral presentation (learning-by-teaching, conceptual integration) Accordingly, this study addressed the following research questions:
- RQ1: Does repetitive handwritten practice improve retention and exam performance in ChatGPT-assisted computational thinking courses?
- RQ2: Does oral presentation enhance conceptual understanding and engagement, even if exam performance is lower?

2. Methodology

2.1. Research Context and Participants

This study was conducted in an undergraduate Introduction to Computer Science (CS1) course at a comprehensive university in Taiwan. The course covered core computational thinking skills, including algorithmic reasoning, control structures, data representation, and basic Python programming, over an 18-week semester. A total of 120 students participated in two intact class sections (Class A and Class B), with 60 students in each section. Both sections were taught by the same instructor using identical syllabi, materials, assignments, and assessments. Because students enrolled before the instructional differences were announced, potential self-selection bias was reduced.

2.2. Research Design

A quasi-experimental design was employed to compare two pedagogical reinforcement strategies in a ChatGPT-assisted learning environment. Random assignment was not feasible due to institutional constraints; therefore, intact class sections were used. To reduce major confounding influences, both sections followed the same course schedule, received the same lectures, and had equal access to ChatGPT as a supplementary learning tool throughout the semester. The only systematic difference between the two groups was the reinforcement strategy embedded in weekly learning activities.

2.3. Instructional Conditions

Class A emphasized repetitive handwritten practice, requiring students to regularly write out solutions to computational problems to support procedural fluency and active recall. Class B emphasized oral presentation, in which students periodically explained computational concepts and applications, thereby promoting conceptual understanding through learning-by-teaching. In both conditions, ChatGPT could be used as a supplementary resource outside formal assessments.

2.4. Assessment and Data Analysis

Both sections completed the same assessments, including a pre-test in Week 1, a midterm exam, and a comprehensive final exam administered without access to ChatGPT. The final exam served as the primary outcome measure of independent computational thinking and problem-solving ability. Group differences were analyzed using independent-samples t-tests for the pre-test, midterm, and final exam. Descriptive statistics were used to summarize score distributions, and post-course survey responses were analyzed descriptively to support interpretation of the quantitative results.

3. Results and Discussion

3.1. Assessment Performance Across the Semester

Table 1 summarizes the performance of Class A and Class B across the three assessments. The pre-test results showed no significant difference between Class A and Class B (69.48 vs. 70.82, $p = 0.78$), indicating comparable baseline knowledge. Likewise, no significant group difference was observed on the midterm exam (76.7 vs. 74.4, $p = 0.80$), suggesting that short-term performance remained similar under equivalent ChatGPT access.

In contrast, a clear difference emerged on the AI-free final exam. Class A scored significantly higher than Class B (52.3 vs. 41.8, $p = 0.01$), with a mean difference of 10.5 points and a medium-to-large effect size (Cohen's $d = 0.70$). Because the final exam was completed without ChatGPT, this result suggests that repetitive handwritten practice more effectively supported procedural fluency and independent problem-solving under AI-free conditions.

Table 1: *Performance Comparison Across Three Assessments*

Measure	*Class A (M ± SD)*	*Class B (M ± SD)*	*p-value*
Pre-test (Week 1)	*69.48 ± 21.77*	*70.82 ± 17.95*	*0.78*
Midterm Exam (Week 9)	*76.7 ± 10.5*	*74.4 ± 10.2*	*0.80*
Final Exam (Week 18)	*52.3 ± 17.1*	*41.8 ± 12.7*	*0.01*

3.2. Distributional Analysis

Beyond mean differences, Fig. 1 further highlights the advantage of handwritten practice on the final exam. As shown in Fig. 1, Class A had a higher median score than Class B (50 vs. 41), a larger proportion of students scoring 60 or above (31.1% vs. 8.8%), and a smaller proportion scoring below 50 (45.9% vs. 63.2%). These patterns suggest that repetitive handwritten practice not only improved final exam performance but also shifted the score distribution upward by reducing the lower tail. This finding indicates that practice-based reinforcement supported more consistent mastery of core computational skills under AI-free assessment conditions.

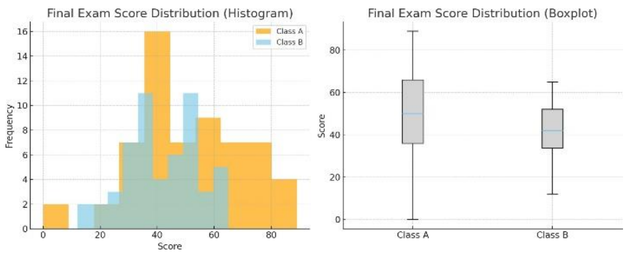


Fig. 1 Final Exam Score Distribution for Class A (Handwritten Practice) and Class B (Oral Presentation). Left: Histogram. Right: Boxplot.

3.3. Qualitative Differences in Exam Performance

Qualitative review of student responses supported the patterns shown in Fig. 1. On step-by-step computational tasks, Class A students more often produced complete and logically coherent solutions, whereas Class B students more frequently demonstrated partial understanding with incomplete procedural execution. On conceptual multiple-choice items, the group difference was smaller, and Class B sometimes performed comparably or slightly better. This pattern suggests that handwritten practice primarily strengthened procedural fluency, while oral presentation supported conceptual articulation.

3.4. Student Perceptions and Engagement

Post-course survey responses further supported these findings. In Class A, 90% of students agreed or strongly agreed that regularly writing out solutions improved their understanding and exam confidence. In Class B, 67% reported that preparing presentations enhanced conceptual clarity and engagement. However, some Class B students also noted difficulty transferring conceptual understanding to independent problem-solving on the final exam. These perceptions reinforce the trade-off between conceptual engagement and procedural mastery in AI-assisted learning environments.

4. Discussion and Conclusions

This study examined how pedagogical reinforcement strategies shaped learning outcomes in a ChatGPT-assisted computational thinking course. Although both groups had equal access to ChatGPT and showed similar performance at baseline and midterm, clear differences emerged on the AI-free final exam. Students in the handwritten-practice group scored significantly higher than those in the presentation group (52.3 vs. 41.8, $p = 0.01$), and distributional results further showed fewer low-performing students and more high-performing students in the practice group. Qualitative and survey findings also suggested that handwritten practice more strongly supported procedural execution and exam confidence, whereas oral presentation was more closely associated with conceptual engagement and verbal articulation.

These findings suggest that generative AI is most effective as a learning scaffold when paired with instructional structures that promote active recall, procedural fluency, and independent problem-solving. While AI-assisted learning may improve short-term performance, durable learning appears to depend more strongly on reinforcement strategy than on AI access alone. In this sense, generative AI functions more as a pedagogical amplifier than as a substitute for effective instructional design.

Several limitations should be noted. First, the quasi-experimental design using intact classes limits strong causal inference. Second, procedural and conceptual learning were inferred from performance patterns and qualitative evidence rather than measured through separate validated instruments. Third, actual ChatGPT usage behaviors were not logged, making it difficult to determine how students differentially relied on AI during the semester. Future research may incorporate AI usage logs and delayed post-tests to further examine long-term learning effects in AI-assisted environments.

References

[1] C. Baek, T. Tate, and M. Warschauer, " 'ChatGPT seems too good to be true': College students' use and perceptions of generative AI," Computers and Education: Artificial Intelligence, vol. 7, Art. no. 100294, 2024.

[2] S. Marwan, G. Gao, S. Fisk, T. W. Price, and T. Barnes, "Adaptive immediate feedback can improve novice programming engagement and intention to persist in computer science," in Proceedings of the 2020 ACM Conference on International Computing Education Research (ICER), Aug. 2020, pp. 194–203.

[3] Z. Ahmed, S. S. Shanto, and A. I. Jony, "Potentiality of generative AI tools in higher education: Evaluating ChatGPT's viability as a teaching assistant for introductory programming courses," STEM Education, vol. 4, no. 3, pp. 165–182, 2024.

[4] C.-H. Chiu and C.-E. Wu, "Practical impact of ChatGPT in an introduction to computer science course: Exam scores and real learning effectiveness," in Proceedings of the 13th International Conference on Information and Education Technology (ICIET), Apr. 2025, pp. 1–6. IEEE.

[5] L. Krupp, S. Steinert, M. Kiefer-Emmanouilidis, K. E. Avila, P. Lukowicz, J. Kuhn, et al., "Unreflected acceptance: Investigating the negative consequences of ChatGPT-assisted problem solving in physics education," arXiv preprint, arXiv:2309.03087, 2023.

[6] O. Dikken, B. Limbu, and M. Specht, "Deliberate practice of handwriting: Supervision under the ghost of an expert," in Proceedings of the European Conference on Technology Enhanced Learning (EC-TEL), Sep. 2022, pp. 434–440. Cham, Switzerland: Springer.

[7] E. Kasneci, K. Seßler, S. Küchemann, M. Bannert, D. Dementieva, F. Fischer, et al., "ChatGPT for good? On opportunities and challenges of large language models for education," Learning and Individual Differences, vol. 103, Art. no. 102274, 2023.

[8] L. Stuchlikova and M. Weis, "From information to insight: Reimagining critical thinking pedagogy in the age of artificial intelligence," in Proceedings of the 2024 International Conference on Emerging eLearning Technologies and Applications (ICETA), Oct. 2024, pp. 591–598. IEEE.

Hsu, T. C., & Ogata, H. (Eds.). (2026). *Proceedings of the 10th International Conference on Artificial Intelligence, Computational Thinking and STEM Education* .Kumamoto: The Asia-Pacific Society for Computers in Education.

Design and Application of STEM Digital Games in Taiwanese kindergartens

Joni Tzuchen TANG[1*]& Pin CHEN[2]
[1,2] National Taiwan University of Science and Technology, Taiwan
[*] jttang0@mail.ntust.edu.tw

Abstract: *Early childhood is a crucial transition period for children as they move from family-based education to formal schooling. This is especially true during the transition from kindergarten to elementary school, where children's learning approaches and abilities face a shift towards more structured learning. Early intervention in STEM (Science, Technology, Engineering, Mathematics) education at the kindergarten stage helps effectively promote children's creativity, logical thinking, and exploratory spirit. These concepts align closely with the design of learning areas in kindergartens. Through open-ended and exploratory methods, learning areas provide cross-disciplinary learning experiences for young children, further fostering STEM development. Digital games, as a highly interactive and engaging learning tool, can stimulate children's autonomous learning and critical thinking through gamified tasks and challenges. This aligns with the exploration and innovation emphasized in STEM education. This study aims to design and develop STEM digital games for young children and explore their application in enhancing children's STEM literacy. The research will adopt a mixed-methods approach, combining quantitative data with qualitative observations, to comprehensively analyze children's learning behaviors and responses during digital gameplay. The research findings will offer concrete game design principles and implementation strategies for educators and provide theoretical and practical guidance for future early childhood education policies and curriculum design. The research project will focus on the learning needs of children in Taiwanese kindergartens and test game design. This research will provide innovative theoretical support and practical evidence for the design of STEM education digital games and validate their effectiveness from multiple perspectives, aiming to enhance young children's STEM literacy and learning motivation.*

Keywords: Early childhood education, STEM education, Digital game-based learning, Kindergarten learning areas, STEM literacy, Learning motivation

1. Introduction

The transition from the final year of kindergarten to elementary school represents a critical developmental stage in which children move from play-based learning toward more structured and goal-oriented educational experiences. During this period, fostering problem-solving abilities, creativity, project-based thinking, and interdisciplinary integration is especially important. In this context, STEM education has been widely recognized as an effective approach for helping children connect disciplinary knowledge with real-world applications and for preparing them for future learning and innovation challenges. This perspective is also consistent with policy initiatives that emphasize the early cultivation of computational thinking and problem-solving skills (Obama, 2013), as well as with previous studies highlighting the importance of STEM education in supporting young children’s STEM literacy development (Chou, 2018). In early childhood settings, the learning-area approach supports exploratory, experiential, and interdisciplinary learning, which aligns closely with the inquiry-oriented nature of STEM education. However, despite this conceptual compatibility, STEM implementation in kindergartens often remains limited in practice, as STEM-related materials are frequently insufficient or fragmented and many educators report a lack of confidence and professional experience in guiding STEM learning activities (Chang, 2001; Chou, 2018; Garbett & Yourn, 2002). In response to these challenges, digital game-based learning has emerged as a promising approach because its interactive, exploratory, and autonomous features can provide young children with engaging and developmentally appropriate opportunities to apply STEM-related concepts through contextualized tasks and problem-solving processes (Tsai & Wu, 2014; Huang & Lin, 2014; Tripp & Liu, 2024). Accordingly, this study aims to design and examine a vehicle-modification digital game for kindergarten children in Taiwan, with the goal of investigating young children's STEM literacy development, evaluating the game's potential to support competencies such as computational thinking, creativity, and problem solving, and exploring whether learning experiences developed through gameplay can be transferred to everyday contexts.

2. Research Methods

2.1. Game Design

The digital game developed in this study is designed for tablet-based interaction to support young children's autonomous STEM learning through exploratory gameplay. At the beginning of the game, children are introduced to different environments, including grassland, rocky terrain, and snowy terrain, through animated scenes that highlight key environmental characteristics. Based on the selected environment, children assemble a vehicle by combining multiple functional components and explore how different configurations influence vehicle performance. After completing the assembly, they operate the vehicle by using simple symbolic commands to accomplish task-based goals, such as coin collection. To encourage reflection and iterative improvement, the game also includes a review mechanism that presents the child's previous vehicle design for comparison, supporting self-reflection and problem-solving during repeated gameplay. Through these simulated environmental challenges and design tasks, the game aims to foster spatial reasoning, logical thinking, foundational engineering concepts, and self-directed learning. The game interface is shown in Figure 1.

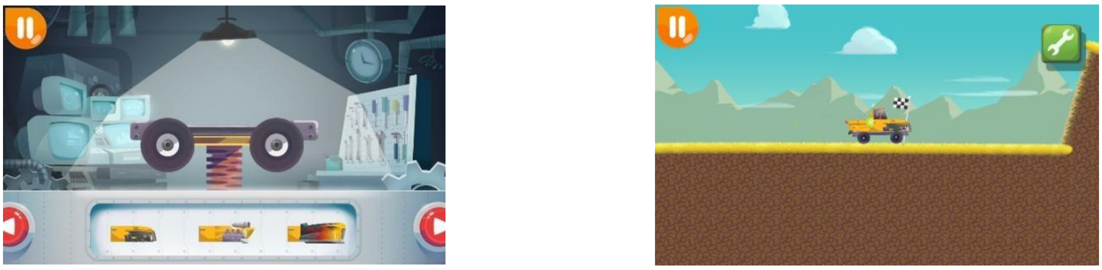


Figure 1. Game screen (Vehicle assembly and successful task completion.)

2.2. Child Observation

This study will be conducted from February to May 2026 in Montessori-related kindergartens in Taiwan, adopting a longitudinal research design to examine the developmental process of

children's STEM literacy. Because STEM literacy is considered an internalized and gradual learning outcome rather than an immediate result, periodic observation across multiple time points is essential for capturing meaningful changes in children's learning behaviors (Tsai & Wu, 2014; Huang & Lin, 2014; Tripp & Liu, 2024). Longitudinal research, which involves systematic and repeated observation of the same participants, is particularly suitable for early childhood studies, as developmental changes often emerge progressively and cannot be fully observed within a short period (Cooper & Schindler, 2003; Tang, 2019).

In this study, children will engage in autonomous digital game-based learning activities on a weekly basis. Both qualitative and quantitative data will be collected through continuous observation to examine children's engagement, problem-solving behaviors, and application of STEM-related concepts during gameplay. Observations will focus on children's interactions with the digital game, their decision-making during vehicle assembly and operation, and changes in strategies and behaviors across repeated gameplay sessions. This longitudinal approach enables the identification of learning patterns and developmental trajectories over time and aligns with established practices in early childhood research, such as large-scale longitudinal studies, making it an appropriate method for examining the internalization of STEM literacy in young children (Dornan et al., 2015; Ho, 2017).

2.3. Curriculum and Instructional Design

The research procedure is illustrated in Figure 2, while the corresponding research schedule and timeline are presented in Figure 3. The study is conducted in four sequential stages to systematically develop, implement, and evaluate a digital game for supporting young children's STEM literacy learning. First, instructional materials and the digital game are designed and developed through iterative refinement, including expert reviews to ensure developmental appropriateness and educational effectiveness. Second, the digital game is implemented in kindergarten settings, where children participate in weekly autonomous gameplay, and data are collected through classroom observations, teacher and parent records, and backend gameplay logs. Third, both quantitative and qualitative analyses are conducted to examine changes in children's learning behaviors, problem-solving strategies, and STEM-related competencies across the intervention period. Finally, based on the results of data analysis, practical design principles and implementation guidelines are proposed to support the integration of digital games into early childhood STEM education and to inform future instructional practice. The duration and overlap of each research phase across the first year are summarized in Figure 3.

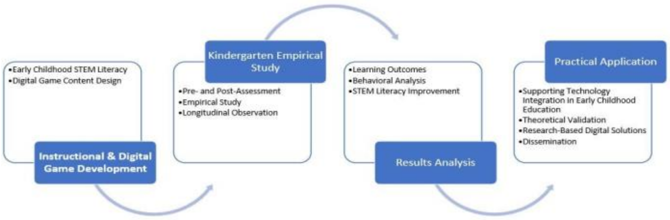


Figure 2. Research process

Figure 3. Research Schedule and Timeline

3. Expected Impact

As global attention to STEM literacy and early childhood educational innovation continues to grow, this study is expected to contribute an innovative instructional perspective to early childhood STEM education by demonstrating the feasibility and educational value of integrating digital game-based learning into kindergarten settings. Through the design and application of a vehicle-modification digital game, the study aims to show how interactive and playful digital environments can support young children's learning motivation, engagement, problem-solving, creativity, and emerging STEM literacy in developmentally appropriate ways. In addition to providing theoretical and empirical support for the use of digital games in early childhood STEM learning, this study is also expected to offer practical design principles and implementation guidance for educators, curriculum designers, and researchers seeking to integrate digital tools into early childhood learning contexts. Moreover, by considering issues related to authentic classroom use, instructional sustainability, and collaboration among teachers, parents, and related partners, the study seeks to bridge research and practice and to promote more engaging, flexible, and meaningful STEM learning experiences for young children.

Acknowledgement

This research was supported by funding from the National Science and Technology Council (NSTC) of Taiwan. The authors would like to express their sincere gratitude for the financial support that made this study possible.

References

Chang, Y. C. (2001). Early childhood curriculum design and learning area implementation (in Chinese).
Taipei, Taiwan: Psychological Publishing.

Chou, S. H. (2018). Early childhood STEM education: Theory and practice (in Chinese). Taipei, Taiwan: Wu-Nan Book Inc.

Cooper, D. R., & Schindler, P. S. (2003). Business research methods (8th ed.). New York, NY: McGraw-Hill.

Dornan, R., Roberts, M., & Jones, S. (2015). Longitudinal approaches in early childhood education research. Early Childhood Research Quarterly, 31, 15–25. https://doi.org/10.1016/j.ecresq.2014.12.003

Garbett, D., & Yourn, B. (2002). Student teacher knowledge: Knowing and understanding subject matter in early childhood education. Teaching and Teacher Education, 18(4), 467–483. https://doi.org/10.1016/S0742-051X(02)00007-3

Ho, Y. L. (2017). Research methods in early childhood education (in Chinese). Taipei, Taiwan: Psychology Press.

Huang, Y. M., & Lin, Y. T. (2014). Digital game-based learning for problem-solving skills. Computers & Education, 71, 92–103. https://doi.org/10.1016/j.compedu.2013.09.017

Obama, B. (2013). Educate to innovate: Preparing every child for success in a technology-driven world.
Washington, DC: The White House.

Tang, T. C. (2019). Longitudinal research methods in early childhood development (in Chinese). Taipei, Taiwan: Psychological Publishing.

Tripp, S., & Liu, M. (2024). Game-based learning and STEM education: Supporting engagement and problem solving. Educational Technology Research and Development, 72(1), 45–63. https://doi.org/10.1007/s11423-023-10245-7

Tsai, C. C., & Wu, Y. T. (2014). Exploring game-based learning for computational thinking development.
Computers & Education, 73, 142–152. https://doi.org/10.1016/j.compedu.2013.12.004

Hsu, T. C., & Ogata, H. (Eds.). (2026). *Proceedings of the 10th International Conference on Artificial Intelligence, Computational Thinking and STEM Education* .Kumamoto: The Asia-Pacific Society for Computers in Education.

Integrating Picture Books as Scaffolding for Introducing Scratch Programming in Elementary Schools

Pin CHEN[1], Joni Tzuchen TANG [2*]
[1,2] National Taiwan University of Science and Technology, Taiwan
* jttang0@mail.ntust.edu.tw

Abstract: *Scratch is widely used in elementary programming education due to its visual, block-based interface; however, introductory Scratch lessons often face challenges related to instructional quality, particularly in helping young learners understand abstract programming concepts and remain meaningfully engaged. This study proposes a picture-book-based instructional approach aimed at improving the quality of introductory Scratch teaching in elementary classrooms. The curriculum integrates storytelling as an instructional scaffold through a castle-building narrative, in which programming concepts are embedded within coherent and purposeful learning contexts. The picture book is structured into ten construction stages, each corresponding to a specific Scratch block category, including events, motion, looks, sound, control, and operators. By aligning story events with Scratch's color-coded block system, the design seeks to support clearer conceptual explanations and more structured learning experiences. The planned instructional intervention will involve approximately 45 elementary school students in grades three and four and will be implemented as a three-hour introductory unit designed for students' first formal exposure to Scratch. Each lesson follows a consistent instructional flow, including guided picture book reading, explicit instruction of Scratch block concepts, teacher demonstration, hands-on programming activities, and brief reflection. As data collection is scheduled to begin in March 2026, this paper focuses on presenting the curriculum design and planned analytical framework. Data will be collected through students' Scratch projects, learning motivation questionnaires, classroom observations, and instructional reflections. This study aims to provide practical design insights for improving instructional quality and pedagogical coherence in elementary Scratch education.*

Keywords: Scratch programming; picture books; elementary education; storytelling; instructional scaffolding; programming motivation

1. Introduction

Scratch has become a widely adopted visual programming platform in elementary education due to its block-based interface and low entry barrier, and prior research has shown its potential to enhance student engagement and participation in primary school settings (Belessova et al., 2024; Resnick et al., 2009). However, many young learners still experience difficulties during their initial exposure to Scratch, particularly in understanding abstract programming logic and maintaining learning motivation. For students in grades three and four, concepts such as sequencing, control, and events may remain challenging when introduced without sufficient contextual and instructional support (Yang & Lin, 2024).

Previous studies have suggested that narrative-based learning environments can function as effective instructional scaffolds for novice learners. Storytelling and picture books provide concrete and meaningful contexts that help learners make sense of abstract concepts while enhancing engagement and emotional involvement (Bruner, 1996; Dietz et al., 2023). In the context of programming education, narrative-based approaches have been found to support children's understanding by situating computational concepts within purposeful and relatable activities.

Building on this perspective, this study proposes a picture-book-based approach to introducing Scratch programming in elementary schools. The curriculum is designed around a castle-building story, in which each construction stage corresponds to a specific Scratch block category. Through this narrative structure, programming concepts are embedded within meaningful tasks, allowing students to associate block colors and functions with story events. Such alignment between instructional materials and learning activities has been identified as an important factor in improving instructional quality in technology-supported learning environments (Kirschner et al., 2006).

The purpose of this study is to explore the potential of picture books as an instructional scaffold for introductory Scratch programming, with particular attention to how narrative integration may contribute to clearer conceptual explanations and more coherent learning experiences. Accordingly, the study addresses the following research questions:

RQ1: How does a picture-book-based approach influence students' engagement and learning motivation in introductory Scratch lessons?

RQ2: How does the narrative-based curriculum support students' understanding of basic Scratch concepts?

RQ3: What design considerations emerge during classroom implementation?

2. Method

Figure 1 illustrates the overall research schedule and methodological design of the study. The planned study involves approximately 45 elementary school students from grades three and four, all of whom are novice Scratch learners. The instructional intervention consists of a three-hour introductory unit designed to support students' first formal exposure to Scratch programming. The curriculum is implemented through a picture book centered on a castle-building narrative, with each lesson following a consistent instructional flow that includes guided story reading, introduction of Scratch block categories, teacher demonstration, student hands-on practice, and brief reflection. The intervention is scheduled to begin in March 2026. Data are collected across the five research phases shown in the figure, including picture book design and development, curriculum design and planning, elementary school experimental study, data analysis, and practical application. Planned data sources include students' Scratch projects, learning motivation questionnaires, classroom observations, and instructional reflections. Quantitative data are analyzed descriptively, while qualitative data are examined using thematic analysis to identify patterns related to student engagement and conceptual understanding.

Figure 1. Research Schedule and Timeline

3. Curriculum Design

The picture-book-based Scratch curriculum is structured around ten castle construction stages, each corresponding to a specific Scratch block category, including events, motion, looks, sound, control, and operators. This staged design introduces programming concepts progressively, allowing students to focus on one conceptual theme at a time.

By mapping story events to Scratch block functions, the curriculum leverages Scratch's color-coded interface to support conceptual understanding. The narrative metaphor positions blocks as tools used to complete construction tasks, helping students associate programming logic with purposeful actions.

Each lesson integrates storytelling with hands-on programming activities, ensuring that narrative elements function as instructional scaffolds rather than distractions. This design aims to reduce cognitive load and support meaningful learning during students' initial programming experiences.

3.1. Picture Book Design

The picture book was designed as an instructional scaffold to support elementary students' initial understanding of Scratch programming concepts. Previous research has shown that storytelling and narrative structures can help young learners organize new information, reduce cognitive load, and increase engagement by embedding abstract concepts within meaningful contexts (Bruner, 1996; Robin, 2008). Drawing on this perspective, the picture book adopts a castle-building narrative in which a king commissions an architect to construct the strongest castle on newly acquired land, providing a coherent and motivating storyline for introducing programming concepts (see Figure 2).

The story is structured into ten construction stages, each representing a distinct phase of castle building and corresponding to a specific Scratch block category. These categories include events, motion, looks, sound, control, and operators. By aligning each stage of the narrative with a single block category, the design allows students to focus on one conceptual theme at a time, supporting gradual knowledge construction. This staged progression reflects principles of instructional scaffolding, in which instructional support is systematically provided and adjusted as learners develop understanding (Wood, Bruner, & Ross, 1976).

To further bridge narrative understanding and programming execution, the picture book integrates story events directly with Scratch-style interaction cues. For example, the initiation of construction work is visually linked to the "when green flag clicked" event block, prompting

learners to associate story actions with program execution logic (see Figure 3). In addition, the picture book leverages Scratch's color-coded block system as a visual–semantic cue. Story events are intentionally associated with block colors and functions, encouraging students to form connections between narrative actions and programming logic. Such visual–narrative alignment has been shown to support comprehension and recall among novice learners, particularly in technology-enhanced learning contexts (Mayer, 2009). Through this design, the picture book serves not only as a motivational element but also as a conceptual bridge between storytelling and computational thinking.

Figure 2. Picture book opening scene introducing the learning context.

Figure 3. Picture book scene supporting programming concept introduction.

3.2. Curriculum and Instructional Design

Building on the picture book design, the instructional curriculum was developed to integrate storytelling with hands-on Scratch programming activities in a structured and pedagogically coherent manner. Research on introductory programming education suggests that novice learners benefit from clearly sequenced instruction, explicit concept explanation, and opportunities for immediate application (Resnick et al., 2009; Lye & Koh, 2014). Accordingly, the curriculum was designed as a three-hour introductory unit that functions as students' first formal exposure to Scratch.

Each lesson follows a consistent instructional flow that aligns narrative elements with programming practice. Lessons begin with guided reading and discussion of a picture book segment, during which the teacher highlights the construction task and introduces the

corresponding Scratch block category. This is followed by explicit instruction and teacher demonstration of block functions, emphasizing their logical roles rather than technical details. Students then engage in hands-on programming tasks that directly mirror the story events, allowing them to apply newly introduced concepts in a meaningful context. Lessons conclude with brief reflection or sharing activities, during which students explain how their programs relate to the narrative.

This instructional flow, illustrated in Figure 4, is intended to support instructional quality by maintaining alignment between story context, conceptual explanation, and learner activity. By systematically connecting narrative understanding with programming practice, the curriculum aims to create a coherent learning experience that supports both engagement and conceptual clarity. Such alignment between instructional design and learning activities has been identified as a key factor in effective technology-supported instruction for young learners (Kirschner, Sweller, & Clark, 2006).

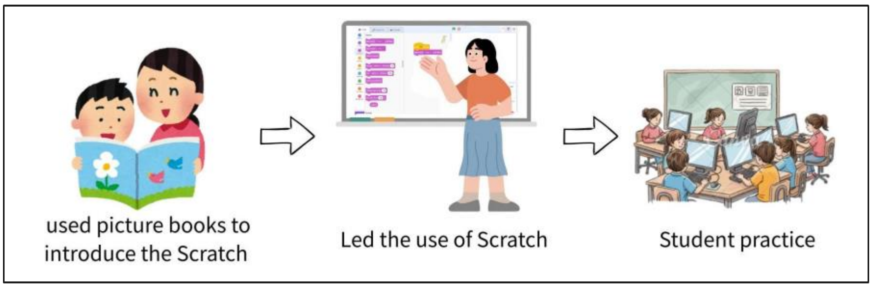


Figure 4. Flowchart for Introducing Picture Books into Scratch Courses

4. Anticipated Results

The study anticipates that students will demonstrate high engagement and positive learning attitudes during the picture-book-based Scratch lessons. The narrative context is expected to support students' willingness to participate in activities and reduce anxiety associated with initial programming experiences.

Students' Scratch projects are expected to reflect basic understanding of introduced block categories, with students demonstrating the ability to apply blocks in ways consistent with story-based tasks. Qualitative observations may reveal that students use narrative references when explaining their programming decisions, indicating the scaffolding role of the picture book.

5. Discussion

This study highlights the potential of picture books as an instructional scaffold for introductory Scratch programming. Embedding programming concepts within a narrative context helps students organize abstract ideas into meaningful structures, supporting both engagement and comprehension.

The curriculum design demonstrates how Scratch's visual and color-coded features can be pedagogically enhanced through storytelling. However, careful instructional pacing and alignment between narrative elements and programming tasks are essential for effective implementation. Future studies may extend this approach to longer interventions or explore its applicability across different age groups.

References

Belessova, D., Ibashova, A., Zhidebayeva, A., Shaimerdenova, G., & Nakhipova, V. (2024). The impact of Scratch on student engagement and academic performance in primary schools. Open Education Studies, 6, 20220228. https://doi.org/10.1515/edu-2022-0228

Bruner, J. (1996). The culture of education. Harvard University Press.

Dietz, G., Tamer, N., Ly, C., Le, J. K., & Landay, J. A. (2023). Visual StoryCoder: A multimodal programming environment for children's creation of stories. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23) (Article 96, 16 pages). ACM. https://doi.org/10.1145/3544548.3580981

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. Educational Psychologist, 41(2), 75–86. https://doi.org/10.1207/s15326985ep4102_1

Lye, S. Y., & Koh, J. H. L. (2014). Review on teaching and learning of computational thinking through programming. Computers in Human Behavior, 41, 51–61. https://doi.org/10.1016/j.chb.2014.09.012

Mayer, R. E. (2009). Multimedia learning (2nd ed.). Cambridge University Press.

Resnick, M., Maloney, J., Monroy-Hernández, A., Rusk, N., Eastmond, E., Brennan, K., Millner, A., Rosenbaum, E., Silver, J., Silverman, B., & Kafai, Y. (2009). Scratch: Programming for all. Communications of the ACM, 52(11), 60–67. https://doi.org/10.1145/1592761.1592779

Robin, B. R. (2008). Digital storytelling: A powerful technology tool for the 21st century classroom.

Theory Into Practice, 47(3), 220–228. https://doi.org/10.1080/00405840802153916

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. Journal of Child Psychology and Psychiatry, 17(2), 89–100. https://doi.org/10.1111/j.1469-7610.1976.tb00381.x

Yang, T.-C., & Lin, Z.-S. (2024). Enhancing elementary school students' computational thinking and programming learning with graphic organizers. Computers & Education, 209, 104962. https://doi.org/10.1016/j.compedu.2023.104962

Academic Judgment in AI-Augmented Learning Environments: The Roles of Cognitive Offloading and AI Literacy

Abstract: *The rapid integration of Generative Artificial Intelligence (Gen-AI) tools—such as ChatGPT, Gemini, and Copilot—into higher education is reshaping how students and academics engage with information, make decisions, and exercise academic judgment. While Gen-AI offers unprecedented opportunities for efficiency, creativity, and cognitive augmentation, an emerging concern is the unintended shift toward cognitive offloading, whereby individuals delegate internal mental processes to external technologies. This extended abstract presents a study that investigates how Gen-AI reliance influences the quality of academic judgment and explores the moderating role of AI literacy in mitigating potential cognitive dependency.*

Grounded in theories of distributed cognition and metacognition, the study proposes that frequent reliance on Gen-AI may unintentionally alter users' cognitive habits. Instead of engaging in active reasoning and critical evaluation, users may adopt a pattern of passive acceptance of AI-generated outputs—a phenomenon we identify as judgment dependency. This shift has implications not only for academic decision-making but also for individuals' confidence in their own critical thinking capabilities. In parallel, the study posits that stronger AI literacy—encompassing understanding of AI's mechanisms, limitations, and biases—may act as a protective factor, enabling users to employ AI as a cognitive support rather than a cognitive substitute.

To examine these dynamics, the study proposes a Sequential Explanatory Mixed-Methods Design. Phase 1 involves a large-scale quantitative survey targeting academics and postgraduate students. Using Partial Least Squares Structural Equation Modelling (PLS-SEM), the study models the relationships between Gen-AI Reliance, Cognitive Offloading, Critical Thinking Self-Efficacy, and Academic Decision Confidence. This phase aims to establish the extent of AI-related cognitive shifts and identify significant predictors of judgment quality.

Phase 2 employs qualitative methods—Think-Aloud Protocols and semi-structured interviews—to provide deeper insight into the cognitive processes underlying observed quantitative patterns. Participants complete evaluative tasks with optional AI assistance while verbalizing their thought processes, allowing researchers to identify moments of cognitive shortcutting, over-reliance, or critical engagement. Data will be analyzed using Reflexive Thematic Analysis to capture nuanced behaviours such as trust calibration, epistemic vigilance, and metacognitive monitoring during AI-mediated judgment tasks.

This study makes several novel contributions. Conceptually, it extends the discourse on AI in education from issues of academic integrity to cognitive integrity, shifting attention from whether students use AI to how AI use reshapes thinking itself. Empirically, it introduces the constructs of cognitive offloading and judgment dependency into the Malaysian higher education context. Practically, it generates evidence-based recommendations for AI-integration guidelines, curriculum design, and AI literacy initiatives that preserve human critical judgment in AI-supported learning environments.

Expected outputs include a validated Conceptual Model of AI-Mediated Academic Cognition, peer-reviewed publications, and policy-oriented guidelines for sustainable AI adoption within higher education institutions. Ultimately, the study broadens the vision of Industry 5.0 by fostering responsible human-machine collaboration where AI amplifies—rather than replaces—human academic intelligence.

Keywords: Generative Artificial Intelligence; Cognitive Offloading; Academic Judgment; AI Literacy; Critical Thinking Self-Efficacy

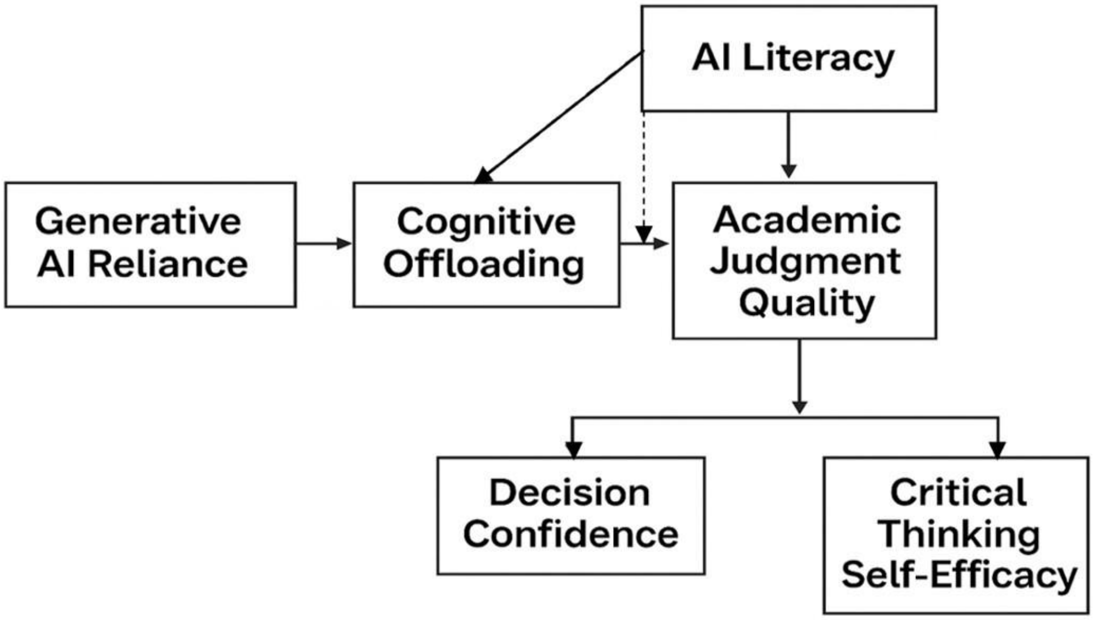


*Fig. 1***.** Conceptual Framework

Hsu, T. C., & Ogata, H. (Eds.). (2026). *Proceedings of the 10th International Conference on Artificial Intelligence, Computational Thinking and STEM Education* .Kumamoto: The Asia-Pacific Society for Computers in Education.

Effects of SRL Strategies and an LLM-RAG Pedagogical Agent on High School Students' Learning Outcomes in AIoT-STEM Hands-on Activities

Hsien-Sheng Hsiao[1], Yu-Hung Chien[2], Guang-Han Chung[3*]

[1,2,3] Department of Technology Application and Human Resource Development, National Taiwan Normal University, Taipei, Taiwan.

[*]gramchung@gmail.com

Abstract: *This study examines the feasibility and pedagogical effects of an instructional model that integrates self-regulated learning (SRL) strategies with a large language model (LLM)–based pedagogical agent enhanced by retrieval-augmented generation (RAG) in a senior high school AIoT–STEM hands-on activities. Although LLMs can support programming learners through debugging assistance and conceptual explanations, their outputs may contain hallucinations or drift beyond the instructional context. To strengthen response reliability and instructional alignment, the proposed system employs a retrieve-then-generate RAG pipeline: relevant excerpts are first retrieved from a teacher-curated instructional knowledge base, and the LLM subsequently generates contextualized guidance with traceable sources.*

This pilot implementation adopts a single-group design and is conducted in one senior high school over a 10-week program (100 minutes per week). The hands-on project centers on an AIoT–STEM "Smart Trash Can" project. Students use an Arduino UNO microcontroller to integrate sensors and actuators and iteratively develop and validate key functions, including proximity-triggered lid opening, fill-level detection, and cloud reporting. Learning outcomes are evaluated using pre–post measures of AIoT–STEM knowledge and programming self-efficacy, complemented by artifact-based performance assessment and backend interaction indicators derived from platform logs. By triangulating learning-process evidence with outcome measures, this study aims to clarify how routinized SRL cycles and RAG-grounded feedback jointly support students' progression in complex programming and hands-on problem solving. The findings are expected to inform actionable design recommendations for classroom deployment, instructionally aligned generative-AI support.

Keywords: AIoT–STEM hands-on activities; pedagogical agent; retrieval-augmented generation; self-regulated learning

1. Introduction

Amid the wave of 21st-century educational reform, rapid technological advances have accelerated instructional transformation. Education increasingly emphasizes integrating information technologies and emerging tools (e.g., generative AI) into learning processes to support students' self-directed learning and improve learning outcomes (Córdova-Esparza, 2025). Self-regulated learning (SRL) foregrounds proactive goal setting, strategic regulation, progress monitoring, and reflective revision, and is critical for sustaining engagement and performance in complex tasks (Zimmerman, 2002).

In secondary education, STEM and programming are widely viewed as key pathways for developing logical reasoning and problem-solving. Programming is increasingly embedded in STEM hands-on activities, where students iteratively design, test, and refine solutions to authentic problems. Prior work suggests that integrating programming into STEM hands-on learning can enhance students' knowledge outcomes (Hsiao et al., 2024) and that embedding programming and AI concepts into core subjects can support broader implementation (Akhmetova et al., 2025).

However, many high school students struggle in early programming learning, which can increase frustration and reduce engagement (Azaiz et al., 2024). These challenges may intensify in STEM hands-on contexts due to concurrent demands of coding, debugging, workflow coordination, and artifact construction under time constraints, while teachers often cannot provide timely, individualized support. Although LLMs (e.g., ChatGPT) can assist with debugging and conceptual understanding (Groothuijsen et al., 2024), they may generate hallucinations and unreliable information, undermining feedback credibility (Huang et al., 2025). Retrieval-Augmented Generation (RAG) can improve response accuracy and consistency by grounding generation in retrieved external knowledge and shows promise for reducing hallucinations in programming education (Wang et al., 2025; Zhang & Zhang, 2025), yet accuracy alone may not ensure sustained engagement or strategic regulation.

Accordingly, three gaps remain: (1) novices often lack systematic debugging strategies and foundational concepts, leading to repeated ineffective approaches and reduced persistence (Azaiz et al., 2024; Huang et al, 2025); (2) classrooms—especially during STEM hands-on work—lack scalable, differentiated support, highlighting the need for SRL-oriented process scaffolding; and (3) LLM hallucinations and citation errors necessitate controllable, traceable knowledge support (Huang et al., 2025).

To address these gaps, this study adopts a learning-process perspective and proposes "SRL strategies × LLM+RAG pedagogical agent" for STEM hands-on activities. RAG connects to a teacher-curated knowledge base to provide contextualized feedback with traceable sources, while the pedagogical agent supports strategy use across planning, monitoring, and reflection to facilitate self-adjustment (Zimmerman, 2002; Chen, 2022). Using pre–post measures and learning-process data, the study will evaluate effects on AIoT–STEM knowledge, programming self-efficacy, and hands-on performance, providing evidence for classroom applications and risk management of generative AI.

2. Materials and Methods

2.1. Research design and context

This study is a pilot experiment adopting a single-group design to examine the effects of implementing an instructional model of "SRL strategies × LLM+RAG pedagogical agent" in a high-school AIoT–STEM hands-on activities on students' learning performance. The hands-on activity centers on a "Smart Trash Can" as an authentic task situated in smart-city governance and public environmental management. Students use an Arduino UNO as the core controller and integrate an ultrasonic sensor (measuring the internal distance to estimate the fill level), an infrared/distance sensor (detecting user-approach events), a servo motor (automatically opening/closing the lid), and LEDs/buzzer (status feedback). Through the implementation, students complete key functions including approach-triggered lid opening, overflow detection and alerts, disposal count logging, and cloud reporting with real-time notifications, thereby progressively developing an integrated understanding of the AIoT-STEM pipeline: Sensing (Input) – Data – Control – Feedback. The course also highlights themes such as public hygiene, environmental sustainability, and urban efficiency, guiding students to connect technological applications with real-world needs to enhance learning motivation and interdisciplinary integration.

2.2. Participants

The participants will be 10th grade students from a high school. The study will implement a 10-week instructional intervention and STEM hands-on activities with an estimated sample size of 40 students. To ensure compliance with research ethics and data-governance requirements, parental consent forms will be obtained and signed prior to the start of the intervention.

2.3. Design of the SRL × LLM+RAG Pedagogical Agent

This study develops an interactive instructional platform in Python (see Figure 1) with a web-based chat window as the main interface. The system logs students' questions, responses, button clicks, submissions, task operations, and learning progress, and stores all records in a MongoDB database as learning-process data. Teachers can upload and manage course workflows and instructional materials through a teacher-facing interface to support classroom management.

The platform also includes a learning monitoring module with fixed-interval push notifications. At scheduled intervals, it reads students' status and event logs from MongoDB (e.g., completed steps and submissions), determines each student's current learning node using rule-based logic, and pushes guidance such as next-step reminders, prompts to revisit materials, and reflection questions. This provides consistent in-class support while allowing teacher intervention when needed.

To improve response accuracy, the system integrates Retrieval-Augmented Generation (RAG). Course materials (e.g., syllabus, handouts, worksheets, sample code, and operation guides) are chunked, embedded, and stored in a Pinecone vector database. Student queries are embedded and matched via semantic retrieval; the retrieved excerpts and the query are then sent to an LLM (e.g., ChatGPT) to generate responses. Through this "retrieve-then-generate" workflow, responses are more closely aligned with the weekly task context and reduce the risk of unsupported generation and off-scope expansion. Throughout the process, the system retains students' interaction data and key learning steps in MongoDB, providing evidence for teachers and researchers to monitor learning status and analyze learning trajectories.

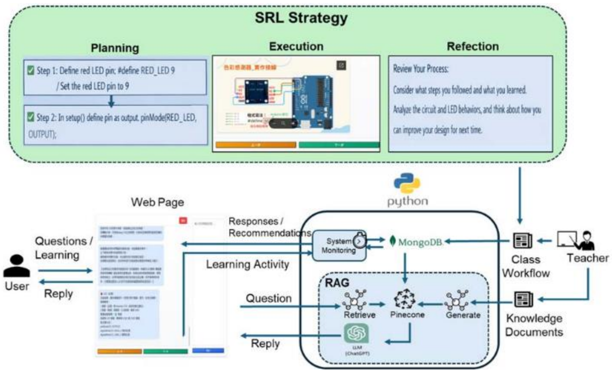


Figure 1. System Architecture Diagram

The SRL scaffolding design in this study is structured around three core processes: planning, monitoring, and reflection. In the planning phase, a linear, predetermined workflow is adopted: the system presents course tasks in a fixed sequence pre-arranged by the teacher, and students are required to complete the steps in order. This arrangement ensures that all students follow the same learning pace and task context, facilitating instructional guidance and stable management of learning trajectories. In the monitoring phase, the system operates through a fixed-interval, reminder-based mechanism: at predefined time intervals, it checks students' current progress and pushes progress reminders or next-step guidance to help students maintain the intended pace and complete the in-class workflow. In the reflection phase, the system prompts students to conduct self-evaluation by asking whether there are still concepts they do not understand, and provides options such as "review again" and "revisit key course materials", supporting students in transforming hands-on experiences into transferable strategic knowledge.

2.4. Instructional Experiment Procedure

This study will recruit first-year students from a senior high school in Taipei City and adopt a pilot single-group design. Under the same syllabus, tasks, and instructional time (100 minutes per week for 10 weeks), the study will examine the effects of implementing an instructional model of "SRL strategies × LLM+RAG pedagogical agent" in an AIoT–STEM hands-on course. In Week 1, a pretest will be administered (cognitive: AIoT–STEM knowledge; affective: programming self-efficacy) without conducting an instructional cycle. Weeks 2–9 constitute the intervention period, following the course progression for programming instruction and the "Smart Trash Bin" project-based hands-on activities. In Week 10, a posttest (same dimensions) will be administered. In addition, both the researcher and the technology education teacher will evaluate the Smart Trash Bin artifacts using the same scoring rubric and consolidate learning-process data. The instructional flowchart is shown in Figure 2.

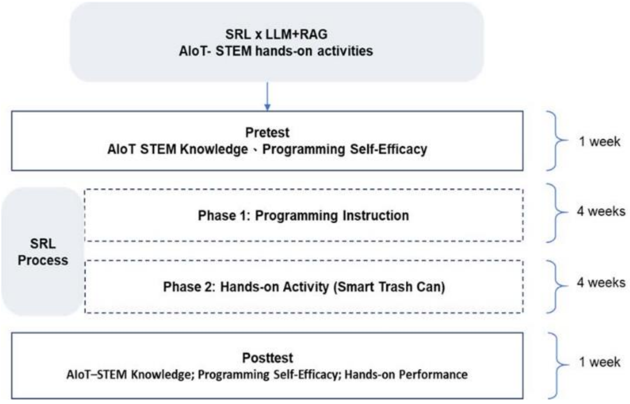


Figure 2. The instructional flowchart

During Weeks 2–9, each class session will follow an SRL cycle of planning–performance–reflection. In the planning phase, students set session goals using system-provided, teacher-defined step-by-step guidance, which serves as the basis for later monitoring and reflection. In the performance phase, students conduct learning activities, coding, and system testing, supported by an LLM+RAG pedagogical agent that provides contextualized prompts, conceptual clarification, and debugging guidance grounded in a teacher-curated knowledge base with traceable sources to reduce hallucination risk. For monitoring, the system applies a fixed time-control rule by triggering "time checks" at preset intervals (e.g., every 10 minutes) and requiring students to report progress and status (e.g., completed steps, sticking points, error messages). In the reflection phase, the system prompts students to evaluate goal attainment based on their work and reports and offers concrete next-step suggestions (e.g., revisiting materials, retrying steps, adjusting the testing sequence). Students must complete each time-check report and submit an end-of-class reflection before proceeding, ensuring a complete and comparable learning trajectory.

2.4. Instructional Experiment Procedure

2.4.1. AIoT–STEM Knowledge Exam

This study will use a researcher-developed AIoT–STEM knowledge Exam to assess students' conceptual understanding of AIoT, including AI concepts, basic IoT knowledge, sensor components, and programming. The test will be reviewed by two experienced technology education teachers to establish content validity. It contains 25 multiple-choice items (total = 100 points) and will be administered as both pretest and posttest (with item/option order rearranged in the posttest). Instructional effects will be examined using ANCOVA, with the posttest score as the dependent variable, group as the independent variable, and the pretest score as the covariate; assumptions will be checked and η^2 will be reported as the effect size.

2.4.2. Programming Self-Efficacy Scale

This study will measure students' confidence in programming using the Computer Programming Self-Efficacy Scale (CPSES) adopted by Hsiao et al. (2023) based on Tsai et al. (2019). The scale includes five dimensions—logical thinking, collaboration, algorithms, control, and debugging—rated on a 6-point Likert scale (1 = strongly disagree to 6 = strongly agree). Reported reliability is high (Cronbach's α = .94 overall; subscales α = .79–.94). Instructional effects will be examined using ANCOVA, with posttest CPSES scores as the dependent variable, group as the independent variable, and pretest scores as the covariate; assumptions will be checked and η^2 will be reported as the effect size.

2.4.3. Hands on performance Scale

This study will assess students' hands-on performance using the Creative Product Assessment Matrix (CPAM) (Besemer & Treffinger, 1981), which includes three dimensions: Novelty, Resolution, and Elaboration and Synthesis (Lin et al., 2019; Hsiao et al., 2023). Hands-on performance will be scored across three dimensions with nine sub-items, each rated 1–5 (total 45 points), with higher scores indicating higher creativity and overall product quality. The rubric will be reviewed by two AIoT–STEM-experienced teachers, and products will be scored by two trained raters with inter-rater reliability (e.g., Cohen's κ) calculated; the average score will be used as the final performance score. Group differences will be tested using independent-samples t-tests, with Cohen's d reported as the effect size.

2.4.4. Interaction Behavior Analysis

This study applies interaction-behavior analysis to capture learning-process evidence that questionnaires and paper-based tests may miss. It profiles students' interaction patterns, help-seeking, and task progression, and examines how the SRL × LLM+RAG pedagogical agent shapes these processes. Using a single-group design, the system follows a retrieve-then-generate workflow, and backend logs provide quantifiable indicators to interpret outcome changes and infer the intervention mechanism.

Backend data include dialogue logs (questions, follow-ups, responses) and operation-event logs (node switching, step navigation, revisiting materials, retries, reflection submissions). These data will be transformed into process indicators covering interaction/questioning (frequency, follow-up ratio, message length), task progression/completion (advancement after feedback, completion rate/time, stalled time), and help-seeking under difficulty (questioning/revisiting/retrying, repeated switching/requests). Descriptive statistics will summarize trends, with additional comparisons by week/task phase or performance level as needed. Finally, process indicators will be linked to outcomes (AIoT–STEM knowledge, programming self-efficacy, hands-on performance) to establish process–outcome connections.

3. Expected Outcomes

- Improved AIoT–STEM knowledge: Students are expected to show pre–post gains after participating in the Smart Trash Can hands-on activities supported by the SRL strategies × LLM+RAG pedagogical agent.
- Enhanced programming self-efficacy: Students' confidence in applying programming concepts and debugging strategies is expected to increase through iterative design-and-test cycles.
- Better hands-on performance: Students are expected to produce higher-quality artifacts and achieve more effective system integration across sensing, control, and feedback functions.
- Richer learning-process evidence: Time-check reports, interaction logs, and key learning-step traces will provide process evidence to explain progress patterns and common bottlenecks under SRL-guided prompts and RAG-grounded feedback.

Acknowledgements

This work was financially supported by the "Institute for Research Excellence in Learning Sciences" of National Taiwan Normal University (NTNU) from The Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan, and sponsored by the National Science and Technology Council(NSTC), Taiwan, R.O.C. under Grant no. 112-2410-H-003 -098 -MY3, 112-2622-H-003 -007 -, 113-2410-H-003 -033 -MY3, 113-2622-H-003 -004 - NSTC 114-2622-H-003-007-.

References

Akhmetova, A.I., Sovetkanova, D.M., Komekbayeva, L.K., Abdrakhmanov, A.E., Yessenuly, D., & Serikova, O.S. (2025). A systematic review of artificial intelligence in high school STEM education research. *Eurasia Journal of Mathematics, Science and Technology Education. 21*(4), em2623.https://doi.org/10.29333/ejmste/16222

Azaiz, I., Kiesler, N., & Strickroth, S. (2024). Feedback-Generation for Programming Exercises With GPT-4. ITiCSE 2024: *Proceedings of the 2024 on Innovation and Technology in Computer Science Education* V. 1, 31–37. https://doi.org/10.1145/3649217.3653594

Chelli, S., Iqbal, U., Hasan, S. M., Atique, S., Singh, O., & Li, Y.-C. J. (2024). Hallucination rates and reference accuracy of ChatGPT and Bard for systematic reviews. *Journal of Medical Internet Research, 26*, e53164. https://doi.org/10.2196/53164

Chen, J. (2022). The effectiveness of self-regulated learning (SRL) interventions on L2 learning achievement, strategy employment and self-efficacy: A meta-analytic study. *Frontiers in Psychology, 13*, 1021101. https://doi.org/10.3389/fpsyg.2022.1021101

Córdova-Esparza, D.-M. (2025). AI-Powered Educational Agents: Opportunities, Innovations, and Ethical Challenges. *Information, 16*(6). https://doi.org/10.3390/info16060469

Groothuijsen, S., van den Beemt, A. A. J., Remmers, J. J. C., & van Meeuwen, L. W. (2024). AI chatbots in programming education: Students' use in a scientific computing course and consequences for learning. Computers and Education: *Artificial Intelligence, 7*, 100290.
https://doi.org/10.1016/j.caeai.2024.100290

Hsiao, H. S., Chen, J. H., Chang, T. L., Li, P. H., & Chung, G. H. (2024). A study on the effects of using the 6E model and a robot teaching assistant on junior high school students' STEM knowledge, learning motivation, and hands-on performance. *Journal of Science Education and Technology, 33*(5), 759-778. https://doi.org/10.1007/s12525-023-00680-1

https://doi.org/10.1609/aaai.v36i11.21557

Huang, C., Fu, L., Hung, S.-C., & Yang, S. (2025). Effect of visual programming instruction on students' flow experience, programming self-efficacy, and sustained willingness to learn. *Journal of Computer Assisted Learning, 41*(1), 1-18. https://doi.org/10.1111/jcal.13110

Wang, Z., Wang, Z., Le, L., Zheng, H.-S., Mishra, S., Perot, V., Zhang, Y., Mattapalli, A., Taly, A., Shang, J., Lee, C.-Y.& Pfister, T. (2025). Speculative RAG: Enhancing Retrieval Augmented Generation through Drafting. Computation and Language. https://doi.org/10.48550/arXiv.2407.08223

Zhang, W., & Zhang, J. (2025). Hallucination Mitigation for Retrieval-Augmented Large Language Models: *A Review. Mathematics, 13*(5), 856. https://doi.org/10.3390/math13050856

Zimmerman, B.J. (2002). Becoming a Self-Regulated Learner: An Overview. *Theory into Practice,41*(2),64–70. https://doi.org/10.1207/s15430421tip4102_2

Hsu, T. C., & Ogata, H. (Eds.). (2026). *Proceedings of the 10th International Conference on Artificial Intelligence, Computational Thinking and STEM Education* .Kumamoto: The Asia-Pacific Society for Computers in Education.

A Theoretical Examination of the Difficulty of Generating Questions in Inquiry-Based Learning and the Prerequisite Role of Computational Thinking

Masanori FUKUI[1*] & Yudai SANO[2]
[1] Graduate School of Engineering, Mie University, Japan
[2] Graduate School of Sciences and Technology for Innovation, Tokushima University, Japan
*fukui@info.mie-u.ac.jp

Abstract: *Inquiry-based learning (IBL) frequently falls short of its intended outcomes, with difficulties often attributed to insufficient motivation or inadequate guidance. This paper focuses on the cognitive prerequisites IBL demands. We argue that "generating a question" and "solving a problem" are cognitively distinct acts and re-examine a four-stage inquiry model to reveal the asymmetry in cognitive demands across stages. Drawing on the distinction between problem formulation and problem solving in computational thinking (CT), we show that IBL tacitly presupposes problem formulation abilities. We extend this analysis to later inquiry stages, propose concrete CT-oriented scaffolding strategies with empirical indicators, and reinterpret why IBL so often fails—not as a motivational deficit, but as a mismatch in cognitive prerequisites. This reinterpretation shifts the locus of explanation from individual learner deficits to the structural demands of inquiry itself.*

Keywords: Computational thinking, problem formulation, problem solving, inquiry-based learning

1. Introduction

Inquiry-based learning (IBL) has long been regarded as a productive form of learning in which students construct knowledge by posing questions, gathering information, and testing hypotheses (Dewey, 1938; Bruner, 1961). Yet a substantial body of research indicates that IBL does not always yield expected outcomes (Kirschner, Sweller, & Clark, 2006): learners often do not know what to investigate, questions remain vague, and inquiries degenerate into formulaic presentations. Hmelo-Silver, Duncan, and Chinn (2007) argued that scaffolding can mitigate these challenges; the present paper asks a prior question—what cognitive prerequisites must be in place before scaffolding can take effect.

Prior work has addressed problem formulation within CT (Wing, 2011; Grover & Pea, 2013) and question generation in IBL (Chin & Osborne, 2008; King, 1992) as separate lines of inquiry. No study has systematically examined how IBL tacitly presupposes the kind of problem formulation that CT foregrounds. In particular, we map the cognitive demands of each inquiry stage onto CT problem formulation, extend this analysis beyond Stage 1 to the investigation and conclusion phases, and propose concrete scaffolding strategies along with falsifiable empirical predictions.

2. Two Cognitive Processes in Inquiry-Based Learning

Banchi and Bell (2008) distinguished four levels of IBL: confirmation, structured, guided, and open inquiry. In confirmation inquiry the question, procedure, and expected result are all provided by the teacher; in open inquiry learners are responsible for all three. The act of posing a question falls increasingly on the learner at higher levels, and it is there that cognitive difficulty concentrates.

We distinguish two core acts. The first act is question generation: determining what to treat as a problem with respect to a given phenomenon—setting boundaries, focusing the investigation, and clarifying constraints in an ill-structured situation. The second is problem solving: reasoning toward a conclusion for an already-formulated question using appropriate methods. Although these acts may appear continuous, the cognitive operations they require differ in kind. King (1992) showed that guided question generation promotes knowledge construction; Jonassen (1997) argued that for novices the greatest difficulty lies in defining the problem itself.

3. A Four-Stage Model of Inquiry and the Asymmetry in Cognitive Demands

A widely shared model describes inquiry as: Stage 1—questioning/problem identification; Stage 2—hypothesis construction; Stage 3—investigation and analysis; Stage 4—conclusion building (NRC, 2000; Pedaste et al., 2015). These stages do not carry equal cognitive demand.

Table 1. *Levels of Inquiry and the Implicit Prerequisite of Problem Formulation*

	Question / Topic	Procedure	Solution / Conclusion	Formulation Agent	CT Prerequisite
Confirmation	Teacher	Teacher	Teacher	Teacher	Low
Structured	Teacher	Teacher	Learner	Mostly teacher	Low
Guided	Teacher	Learner	Learner	Shared	Moderate
Open	Learner	Learner	Learner	Learner	High

Stage 1 is the most cognitively demanding as it requires judging what to constitute as a problem. The CT-prerequisite analysis also extends to later stages. CT prerequisites are not confined to Stage 1; this point is developed in Section 6.1.

4. The Dual Structure within Computational Thinking

Computational thinking (CT) extends beyond algorithmic thinking. Wing (2006) defined CT as the thought process of formulating a problem and its solution so that it can be executed by an information-processing agent, encompassing decomposition, abstraction, and algorithm design. Grover and Pea (2013) further identified pattern recognition as a core CT component, and positioned problem formulation as a precondition for problem solving.

Grover and Pea (2013) positioned problem formulation as a precondition for problem solving, and the NRC (2010) characterized CT as a mode of thinking in which complex problems are defined and structured into solvable units. We distinguish problem formulation—deciding what constitutes a problem—from general problem solving—deciding how to solve a given problem—because the difficulty of IBL concentrates precisely at the formulation phase.

5. The Correspondence between IBL and CT

Stage 1 of inquiry—question setting—can be understood as CT-oriented problem formulation. Table 2 maps six cognitive operations shared between IBL question generation and CT problem formulation. Aho (2012) highlighted reformulation as central to CT. As the level of inquiry increases, formulation responsibility shifts from teacher to learner (Table 1). This asymmetry creates a structural problem that has been largely overlooked. Many policy documents position CT as an ability to be developed through inquiry (ISTE & CSTA, 2011; Weintrop et al., 2016)—emphasizing "inquiry cultivates CT." The reverse dependency has received little attention, and the result is a circular difficulty: CT cannot develop because inquiry never gets started; inquiry cannot get started because CT is lacking.

Table 2. *Shared Cognitive Operations between CT Problem Formulation and IBL Question Generation*

Cognitive Operation	Problem Formulation in CT	Question Generation in IBL
Decomposition	Partition complex problem into manageable sub-problems (Wing, 2006)	Narrow a broad topic into sub-questions of investigable scope
Abstraction	Extract essential features; discard irrelevant detail (Wing, 2006)	Select relevant variables from a phenomenon; set aside the rest

Boundary setting	Define the problem space (Jonassen, 1997)	Determine scope: what will and will not be investigated
Constraint identification	Specify conditions a solution must satisfy (Wing, 2011)	Identify conditions under which a question becomes answerable
Reformulation	Transform problem into a tractable form (Aho, 2012)	Reshape a vague interest into a researchable question
Pattern recognition	Recognize analogous problem structures (Grover & Pea, 2013)	Draw on familiar inquiry patterns to identify an investigative strategy

6. Discussion

The difficulty of IBL is understood here as a mismatch between the degree of problem formulation presupposed before inquiry begins and learners' CT-related abilities—greatest in open and guided inquiry. This perspective redescribes familiar phenomena— "no questions emerge," "the investigation goes in circles"—not as motivational deficits, but as failures to meet cognitive prerequisites.

6.1. CT Prerequisites Beyond Stage 1

In Stage 3, abstraction and pattern recognition are required to select relevant data and apply suitable analytical frameworks. Without the operation of abstraction, learners tend to collect data without any principled basis for selection. In Stage 4, reformulation and constraint identification are required to judge which conclusions are defensible. Chin and Osborne (2008) noted that learners' questions cluster around factual types, rarely reaching hypothesis-forming questions. Within the present framework, this pattern can be interpreted as reflecting a deficit in reformulation ability—an operation required not only at Stage 1 but also when drawing conclusions at Stage 4.

6.2. Scaffolding Strategies

Three practical strategies follow from this analysis. First, teachers should select inquiry levels aligned with learners' formulation capacity. Asking learners to write a one-sentence "what I will and will not investigate" statement before inquiry begins serves as a practical indicator of boundary-setting ability.

Second, a pre-inquiry problem-structuring activity can operationalize three key operations. In a social studies classroom, with "environmental problems" as the chosen topic, the teacher asks: (a) decomposition—"List specific sub-issues (waste, air quality, water)"; (b) constraint identification—"Which sub-issue can you observe or measure given available time and resources?"; (c) reformulation—"Turn your chosen sub-issue into a researchable question (e.g., Does sorting household waste reduce landfill in our city?)". This three-step scaffold applies in any non-computing classroom because it targets generic problem-structuring operations rather than programming knowledge.

Third, when inquiry fails, reframe it as a structural question about absent prerequisites rather than learner attitude or teacher competence.

6.3. Empirical Indicators for Future Validation

Although theoretical, the framework yields testable predictions. If CT-prerequisite mismatch drives inquiry failure, learners who can articulate boundary-setting statements before inquiry should generate more investigable questions than those who cannot. Moreover, structured scaffolding for decomposition and constraint identification at Stage 1 should reduce stalled inquiries—an effect that should be strongest for open inquiry. Coding learner-generated questions in terms of formulation operations (Fukui et al., 2025) offers one practical path toward empirically testing these claims.

7. Limitations and Future Work

Three limitations constrain the current analysis. First, fine-grained analysis of specific cognitive sub-operations remains incomplete; second, the framework has not been empirically tested; and third, individual differences in CT development have not been addressed. Future work should test the scaffolding strategies in Section 6.2 using modified problem-posing methodology (Fukui et al., 2025).

References

Aho, A. V. (2012). Computation and computational thinking. *The Computer Journal*, *55*(7), 832–835. Banchi, H., & Bell, R. (2008). The many levels of inquiry. *Science and Children*, *46*(2), 26–29.

Bruner, J. S. (1961). The act of discovery. *Harvard Educational Review*, *31*(1), 21–32.

Chin, C., & Osborne, J. (2008). Students' questions: A potential resource for teaching and learning science. *Studies in Science Education*, 44(1), 1–39.

Dewey, J. (1938). *Experience and Education*. New York: Macmillan.

Fukui, M., Miyadera, R., Sasaki, Y., & Hirashima, T. (2025). A Preliminary Approach to Quantitative Evaluation of Modified Problem-Posing for Problem Structure Understanding in Computational Thinking. *In Proceedings of the 9th International Conference on Computational Thinking and STEM Education in the Era of AI 2025.*

Grover, S., & Pea, R. (2013). Computational thinking in K–12: A review of the state of the field.

Educational Researcher, *42*(1), 38–43.

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist*, *42*(2), 99–107.

ISTE & CSTA. (2011). *Operational definition of computational thinking for K–12 education*.

International Society for Technology in Education and Computer Science Teachers Association. https://cdn.iste.org/www-root/Computational_Thinking_Operational_Definition_ISTE.pdf

Jonassen, D. H. (1997). Instructional design models for well-structured and ill-structured problem-solving learning outcomes. *Educational Technology Research and Development*, *45*(1), 65–94.

Jonassen, D. H. (2000). Toward a design theory of problem solving. *Educational Technology Research and Development*, *48*(4), 63–85.

King, A. (1992). Facilitating elaborative learning through guided student-generated questioning.

Educational Psychologist, *27*(1), 111–119.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, *41*(2), 75–86.

National Research Council. (2000). *Inquiry and the National Science Education Standards: A guide for teaching and learning*. Washington, DC: National Academies Press.

National Research Council. (2010). *Report of a Workshop on the Scope and Nature of Computational Thinking*. Washington, DC: National Academies Press.

Pedaste, M., Mäeots, M., Siiman, L. A., de Jong, T., van Riesen, S. A. N., Kamp, E. T., Manoli, C. C., Zacharia, Z. C., & Tsourlidaki, E. (2015). Phases of inquiry-based learning: Definitions and the inquiry cycle. *Educational Research Review*, *14*, 47–61.

Weintrop, D., Beheshti, E., Horn, M., Orton, K., Jona, K., Trouille, L., & Wilensky, U. (2016). Defining computational thinking for mathematics and science classrooms. *Journal of Science Education and Technology*, *25*(1), 127–147.

Wing, J. M. (2006). Computational thinking. *Communications of the ACM*, *49*(3), 33–35.

Wing, J. M. (2011). *Research notebook: Computational thinking—What and why?* The Link Magazine, Carnegie Mellon University.

Leveraging AI Feedback within Self-Regulated Learning to Support Algorithmic Problem-Solving: Prior Ability Differences

Zeng-Wei HONG[1*], Ming-Chi LIU[2], Wai-Khuen CHENG[3] & Yu-Lun CHENG[4]
[1,2,4] Department of Information Engineering and Computer Science, Feng Chia University, Taichung City, Taiwan
[3]Department of Computing Faculty of Science, Management & Computing, Universiti Teknologi PETRONAS (UTP), Seri Iskandar, Perak, Malaysia
*zwhong@fcu.edu.tw

Abstract: *Algorithmic problem-solving is a core component of computational thinking (CT) that requires both domain knowledge and effective self-regulated learning (SRL). This study examined how AI feedback supports algorithmic problem-solving within an SRL framework, with particular attention to learners' prior ability. A pretest–posttest design was used to evaluate changes in algorithmic thinking performance, and SRL measures were collected to examine regulatory processes. The results showed that AI feedback significantly improved learners' algorithmic thinking performance. However, its effectiveness varied by prior ability, with significant gains observed only among learners with moderate prior knowledge. In terms of SRL, AI feedback was more strongly associated with in-process regulation (e.g., metacognitive monitoring and effort regulation) than with forethought-related processes. These findings suggest that AI feedback functions primarily as a performance-phase scaffold and highlight the importance of aligning AI-supported feedback with learners' ability levels.*

Keywords: AI feedback, self-regulated learning, computational thinking, algorithmic problem-solving, prior ability

1. Introduction

Algorithmic problem-solving is a core component of computational thinking (CT) in computer science education. It involves processes such as decomposition, abstraction, and algorithm design, which enable learners to construct structured solutions to complex problems (Wing, 2006; Shute et al., 2017). Because such tasks are cognitively demanding, learners' performance depends not only on domain knowledge but also on their ability to regulate their learning processes.

Self-regulated learning (SRL) conceptualizes learning as a cyclical process involving planning, performance, and reflection (Zimmerman, 2002). Within this framework, feedback plays a critical role in supporting learners' monitoring and strategy refinement during problem-solving. In algorithmic tasks, feedback can guide learners to examine their reasoning and iteratively improve their solutions.

Recent advances in generative artificial intelligence have introduced AI feedback as a form of real-time instructional support. By providing prompts and guidance during problem-solving, AI feedback can be integrated into learners' regulatory processes. However, existing studies have primarily focused on overall learning outcomes, with limited attention to how AI feedback functions within SRL processes. In addition, learners' prior ability may influence how feedback is interpreted and utilized, yet empirical evidence in AI-supported contexts remains limited.

Accordingly, this study investigates how AI feedback supports algorithmic problem-solving within an SRL framework, with particular attention to differences associated with learners' prior ability. The following research questions guide this study:

- Does AI feedback facilitate learners' performance in algorithmic problem-solving tasks?
- Do learners with different levels of prior ability demonstrate differences in learning outcomes when using AI feedback?
- How does AI feedback influence learners' self-regulated learning (SRL) during the problem-solving process?

2. Method

2.1. Participants and Design

A total of 25 undergraduate students majoring in computer science and information engineering at a university in Taiwan participated in this study. All participants had prior experience with basic programming and algorithmic concepts (M age = 19.49, SD = 0.51; 16 men and 9 women). This study adopted a single-group design in which all participants completed the same algorithmic problem-solving tasks with AI-generated feedback. Participants interacted individually with ChatGPT to obtain feedback, and all interaction records were collected for analysis.

The study was conducted over five weeks. In Week 1, participants attended an orientation session and completed the pretest. During Weeks 2–4, participants completed one problem-solving task per week with AI feedback. In Week 5, participants completed the post-tests.

2.2. Algorithmic Problem-Solving Tasks and AI-Supported Learning Design

Three CT-oriented algorithmic problem-solving tasks were designed in real-world contexts, focusing on route planning, search strategy, and scheduling. Each task was completed over one week. The tasks were open-ended yet structured, requiring learners to analyze constraints, construct solution steps, and justify their reasoning.

An AI-supported feedback mechanism was integrated into the learning process. After submitting their solutions, learners received AI-generated prompts that encouraged iterative refinement. The AI system acted as a learning assistant by paraphrasing learners' ideas, posing guiding questions, and providing suggestions, rather than supplying direct answers. Through this design, AI feedback supported learners' ongoing monitoring and strategy adjustment during problem-solving (see Figure 1).

Figure 1. Screenshot of the Developed AI Feedback Platform

2.3. Instruments

Algorithmic thinking ability was assessed using the Algorithmic Thinking Test for Adults (ATTA) (Lafuente Martínez et al., 2022), administered as both a pretest and post-test. The instrument demonstrated acceptable internal consistency (Cronbach's α = .854).

Self-regulated learning was measured using selected subscales of the Online Self-Regulated Learning Questionnaire (SRL-O) (Broadbent et al., 2023), including self-efficacy, motivation, planning, metacognition, and effort regulation. The questionnaire was administered as a post-test to examine learners' SRL characteristics during the intervention (Cronbach's α = .944).

3. Results

3.1. Effects of AI Feedback on Algorithmic Thinking Ability

To examine changes in students' algorithmic thinking ability, pretest and post-test ATTA scores were compared using the Wilcoxon signed-rank test. As shown in Table 1, post-test scores were significantly higher than pretest score, with a moderate effect size (r = .40), indicating improvement in algorithmic thinking following AI-supported feedback.

To further examine whether the effects of AI feedback varied by prior ability, participants were classified into high-, moderate-, and low-ability groups based on their pretest ATTA scores, with the top and bottom 27% forming the high- and low-ability groups, respectively, and the remaining participants assigned to the moderate-ability group. As shown in Table 2, only students with moderate prior algorithmic thinking ability demonstrated a significant improvement in ATTA scores after receiving AI feedback, with a moderate effect size (r = .40). No significant pretest–post-test differences were observed for the high- or low-ability groups. These results indicate that the learning benefits of AI feedback were not uniform across ability levels but were most evident among learners with intermediate prior algorithmic knowledge.

Table 1. *Pretest–Post-test Comparisons of ATTA Scores across Prior Ability Levels*

Measure		N	Pretest		Posttest		Z	*p*
			M	SD	M	SD		
ATTA	Overall	25	12.2	2.48	13.2	2.78	2.04*	.045
(Prior Ability)								
	High	6	15.67	1.63	15.67	1.37	0.13	.891
	Mid	13	11.85	0.80	13.23	2.17	2.21*	.027
	Low	6	9.33	0.82	10.67	3.01	0.74	.458

$^*p < .05$

3.2. Descriptive Results of Self-Regulated Learning

Post-test SRL-O results indicated moderate to moderately high levels of self-regulated learning across dimensions. Learners reported relatively higher levels of intrinsic motivation (M = 5.14, SD = 0.98), metacognition (M = 4.92, SD = 0.99), and effort regulation (M = 4.70, SD = 1.07), with slightly lower levels of planning and time management (M = 4.31, SD = 1.17). Overall, the results suggest a tendency toward stronger in-process regulation than forethought-oriented regulation.

4. Discussion and Conclusion

This study examined how AI feedback supports algorithmic problem-solving within an SRL framework, with particular attention to learners' prior ability. The findings indicate that AI

feedback can support improvements in algorithmic thinking performance, suggesting its role as a scaffold in complex problem-solving tasks. However, its effectiveness was not uniform across learners. Improvements were primarily observed among learners with moderate prior ability, suggesting that the benefits of AI feedback may depend on learners' readiness to interpret and apply feedback.

From an SRL perspective, the findings suggest that AI feedback is more closely associated with in-process regulation, such as metacognitive monitoring and effort regulation, than with forethought-related processes. This pattern indicates that AI feedback may function primarily as a performance-phase scaffold, supporting learners during task execution rather than influencing planning behaviors within a short-term intervention.

These findings highlight the importance of aligning AI-supported feedback with learners' prior ability levels when designing algorithmic learning environments. Rather than assuming uniform effectiveness, AI feedback should be considered as a differentiated support mechanism that may benefit specific learner groups.

This study is exploratory in nature and should be interpreted with caution due to the small sample size and the absence of a control group. Future research may further examine how different feedback designs or hybrid feedback approaches can support learners across varying ability levels.

Acknowledgements

This work was supported by the National Science and Technology Council, Taiwan [grant number: NSTC 114-2410-H-035-007 and 114-2410-H-035-005-MY2]

References

Broadbent, J., Panadero, E., Lodge, J. M., & Fuller-Tyszkiewicz, M. (2023). The self-regulation for learning online (SRL-O) questionnaire. *Metacognition and Learning, 18*(1), 135-163.

Hadwin, A. F., Järvelä, S., & Miller, M. (2011). Self-regulated, co-regulated, and socially shared regulation of learning. In B. J. Zimmerman & D. H. Schunk (Eds.), Handbook of self-regulation of learning and performance (pp. 65–84). Routledge.

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, 102274.

Lafuente Martínez, M., Lévêque, O., Benítez, I., Hardebolle, C., & Zufferey, J. D. (2022). Assessing computational thinking: Development and validation of the algorithmic thinking test for adults. *Journal of Educational Computing Research, 60*(6), 1436-1463.

Narciss, S. (2013). Designing and evaluating tutoring feedback strategies for digital learning. Digital *Education Review, 23*, 7-26.

Ng, D. T. K., Tan, C. W., & Leung, J. K. L. (2024). Empowering student self-regulated learning and science education through ChatGPT: A pioneering pilot study. *British Journal of Educational Technology, 55*(4), 1328–1353.

Nicol, D. (2020). The power of internal feedback: exploiting natural comparison processes. *Assessment & Evaluation in Higher Education, 46*(5), 756–778.

Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology, 8*, 422.

Shute, V. J., Sun, C., & Asbell-Clarke, J. (2017). Demystifying computational thinking. Educational research review, 22, 142-158.

Wing, J. M. (2006). Computational thinking. *Communications of the ACM, 49*(3), 33-35.

Theobald, M., & Bellhäuser, H. (2022). How am I going and where to next? Elaborated online feedback improves university students' self-regulated learning and performance. *The Internet and Higher Education, 55*, 100872.

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory into Practice, 41*(2), 64-70.

Hsu, T. C., & Ogata, H. (Eds.). (2026). *Proceedings of the 10th International Conference on Artificial Intelligence, Computational Thinking and STEM Education* .Kumamoto: The Asia-Pacific Society for Computers in Education.

Effects of AI Learning Companions on Learning Performance and Computational Thinking: An ICAP-Based Quasi-Experimental Study

Ting-Han Liao[1], Tsun-Sheng Fan[2], Ting-Chia Hsu[3], Yi-Wen Liao[4*]

[1, 4]Department of Intelligent Commerce, National Kaohsiung University of Science and Technology, Taiwan

[2] Practicum Office, St.Peter Senior High School, Taiwan

[3] Department of Technology Application and Human Resource Development, National Taiwan Normal University, Taiwan

* pinkwen923@gmail.com

Abstract: *AI learning companions are widely considered tools for fostering higher-order cognition and problem-solving; however, empirical evidence in authentic classrooms remains mixed. This study used a quasi-experimental design to examine three conditions—no AI, an AI learning companion, and a non-companion AI tool—on students' learning performance and computational thinking in a Tower of Hanoi unit. Participants were tenth-grade students using Taiwan's adaptive learning platform. Learning outcomes were measured using a pretest–posttest design and analyzed with ANCOVA, while AI interaction data were examined through the ICAP framework to assess cognitive engagement. Results showed no significant differences among conditions after controlling for prior ability. However, qualitative findings revealed that engagement was mainly passive and active, with limited constructive and interactive behaviors. These findings suggest that the effectiveness of AI-supported learning depends less on the presence of AI and more on its ability to promote higher-level cognitive engagement, offering implications for the design of AI learning environments.*

Keywords: AI learning companions; Computational thinking; Cognitive engagement; ICAP framework; the Tower of Hanoi

1. Introduction

With the rapid development of Artificial Intelligence (AI) in education, AI-assisted learning has become a key focus, particularly for its role in supporting individualized adjustments and strategy revision through real-time feedback (Ouyang & Jiao, 2021; Wu & Yu, 2024). In problem-solving contexts, AI provides prompts and feedback for multi-step planning; however, its effects on learning outcomes are inconsistent and depend on task characteristics, instructional design, and the level of learning activities elicited (Sailer et al., 2024; Wu & Yu, 2024), indicating that AI alone does not guarantee improvement. Consequently, research has shifted to examining cognitive engagement as a key mechanism linking instruction and outcomes. The ICAP framework identifies four levels—passive, active, constructive, and interactive—with different learning effects (Chi & Wylie, 2014), and has been widely used to analyze learning behaviors in technology-enhanced environments (Chi et al., 2018). Despite this, empirical evidence remains limited on whether different AI supports in authentic classrooms can enhance higher-level engagement, problem-solving, and computational thinking. Therefore, this study adopted a quasi-experimental design comparing no AI, AI as a learning partner, and non-partner AI tools, examining students' learning outcomes and computational thinking in a Tower of Hanoi unit and analyzing cognitive engagement using ICAP.

2. Methodology

2.1. Research Design

This study employed a quasi-experimental pretest–posttest design to investigate the effects of different AI-assisted learning contexts on students' learning outcomes and computational thinking. Three groups were included: a control group without AI (Group A), an experimental group using the e-Du AI learning partner (Group B), and another experimental group using non–e-Du AI tools (Group C). All groups followed the same instructional schedule and content to control for instructional variables.

Students completed a computational thinking scale and a Tower of Hanoi algorithm test before and after the intervention. In addition, AI usage log data were collected to support qualitative analysis of students' cognitive engagement during the learning process.

2.2. Participants

Participants were three Grade 10 classes from a senior high school, assigned to three groups: Group A (n = 33), Group B (n = 31), and Group C (n = 21). All students had prior experience with digital learning platforms but had not previously learned the Tower of Hanoi unit.

2.3. Learning Materials and AI Interventions

The instructional content focused on the Tower of Hanoi problem, including rules, strategy, and problem-solving processes. All students used the same Adaptive Learning Platform and completed identical tasks. The difference lay in AI support: the control group used no AI features; Experimental Group 1 used the e-Du AI learning partner with real-time feedback and guidance; Experimental Group 2 used AI tools providing hints or explanations without interactive dialogue.

2.4. Instruments

This study utilized a Computational Thinking Scale and a Tower of Hanoi Algorithm Test, both administered before and after the intervention to measure learning outcomes and computational thinking. Computational thinking was further analyzed across five dimensions: problem decomposition, pattern recognition, abstraction, algorithm design, and debugging. Additionally, AI usage logs were collected to record students' interactions with AI tools, including questions, responses, and subsequent actions, serving as data for qualitative analysis.

2.5. Data Analysis

Quantitative data were analyzed using ANCOVA with pretest scores as covariates to examine differences in learning outcomes and computational thinking among groups. Qualitative analysis was conducted based on the ICAP framework, categorizing learning behaviors into Passive, Active, Constructive, and Interactive levels using AI interaction logs to explore students' cognitive engagement patterns

3. Results

3.1. Learning Performance on the Tower of Hanoi Algorithm

In this study, posttest scores on the Tower of Hanoi algorithm test were treated as the dependent variable, and pretest scores were used as the covariate. An analysis of covariance (ANCOVA) was conducted to examine the effects of different AI usage conditions on students' learning outcomes. Table 1 presents the analysis results for the control group (Group A), Experimental Group 1 using the e-Du AI learning partner (Group B), and Experimental Group 2 using non–e-Du AI-assisted tools (Group C).

Table 1. *ANCOVA results for the Tower of Hanoi algorithm posttest*

Source	**Type III SS**	**df**	**MS**	**F**	***p***
PreTest (Covariate)	2779.87	1	2779.87	8.29	.005
Group	191.63	2	95.82	0.29	.752
Error	27150.66	81	335.19		

The ANCOVA results indicated that pretest scores had a significant effect on posttest scores for the Tower of Hanoi algorithm test, suggesting that students' prior ability was associated with their subsequent learning performance. However, after controlling for pretest scores, no statistically significant differences were found among the groups in posttest performance on the Tower of Hanoi algorithm test.

4. Conclusions

This study highlights a paradox in STEM project performance: although university students possess greater technical knowledge, their output is hindered by stress, academic overload, and part-time work. In contrast, school students—despite their lower technical expertise—achieve higher scores due to structured support and focused learning environments.

Quantitative results show school teams averaged 73.1 on a standardized rubric, outperforming university teams, who averaged 65.1. This superiority extended across various criteria, including documentation and project quality. Qualitative data revealed that school students displayed higher motivation and engagement, while university students faced fragmented schedules and external demands.

From a sociological lens, the lower performance of university students is linked to systemic issues in higher education, where survival often trumps creativity. Conversely, school students benefit from consistent guidance and fewer external pressures. This disparity is reinforced by stress data: 75% of university students reported high stress, versus 40% of school students, largely due to lack of mentorship and time constraints.

The findings call for targeted policy changes in higher education to overcome structural barriers impacting STEM project performance. Key recommendations include reducing academic overload through interdisciplinary projects, strengthening mentorship to mirror school-level support, offering financial aid for students with part-time jobs, and promoting stress management via flexible deadlines and mental health services. These measures aim to foster an environment where students can effectively apply their technical skills to real-world challenges, enhancing overall STEM learning outcomes.

References

Czerniewicz, L., & Brown, C. (2014). The habitus and technological practices of rural students: a case study. *South African Journal of Education, 34*(1).

Hernández-Sampieri, R. (2010). *Fundamentos de Metodología de la Investigación*. McGraw-Hill Interamericana de España S.L. https://books.google.com.pe/books?id=WILJNAAACAAJ

Hsueh, S.-L., & Kuo, C.-H. (2016). Factors in an interdisciplinary curriculum for the students of industrial design designing multifunctional products. *Eurasia Journal of Mathematics, Science and Technology Education*, *12*(4), 1075–1089.

Jara, C. A., Candelas, F. A., Puente, S. T., & Torres, F. (2011). Hands-on experiences of undergraduate students in Automatics and Robotics using a virtual and remote laboratory. *Computers & Education*, *57*(4), 2451–2461.

Kim, C., Kim, D., Yuan, J., Hill, R. B., Doshi, P., & Thai, C. N. (2015). Robotics to promote elementary education pre-service teachers' STEM engagement, learning, and teaching. *Computers & Education*, *91*, 14–31.

Levy, S., & Schady, N. (2013). Latin America's social policy challenge: Education, social insurance, redistribution. *Journal of Economic Perspectives*, *27*(2), 193–218.

Montés, N., Zapatera, A., Ruiz, F., Zuccato, L., Rainero, S., Zanetti, A., Gallon, K., Pacheco, G., Mancuso, A., & Kofteros, A. (2023). A novel methodology to develop STEAM projects according to national curricula. *Education Sciences*, *13*(2), 169.

Sampieri, R. H., Collado, C. F., & Lucio, P. B. (1998). *Metodología de la investigación*. McGraw-Hill. https://books.google.com.pe/books?id=e4OnuQAACAAJ

Soto, D. A. (2020). *Technology and the future of work in emerging economies: What is different*.

Hsu, T. C., & Ogata, H. (Eds.). (2026). *Proceedings of the 10th International Conference on Artificial Intelligence, Computational Thinking and STEM Education* .Kumamoto: The Asia-Pacific Society for Computers in Education.

Overcoming Traditional Podcast Search Limitations via RAG: An Investigation into Response Quality and Stability

Kai Chun CHENG[1*],Chia Hao CHIU[2*]
[1,2] Department of Data Science,Soochow University,Taiwan
* tone31369@gmail.com*starklab2020@scu.edu.tw

Abstract: *To address the limitations of traditional keyword-based podcast search systems, which often fail to retrieve relevant content from unindexed audio segments, this study proposes a Retrieval-Augmented Generation (RAG) system. The framework integrates OpenAI Whisper for precise transcription, the bge-m3 model for semantic embeddings, and FAISS for high-precision vector retrieval. The system's performance was rigorously evaluated against a legacy keyword-search platform (Apple Podcasts) based on 600 item responses collected from 30 questionnaires across "Specified Response Questions" and "Spontaneous Inquiries". The results demonstrate that the proposed RAG system ($M = 6.06$, $SD = 0.86$) significantly outperforms the legacy system ($M = 1.34$, $SD = 0.66$), with $p < 0.001$, providing far more accurate and contextually relevant responses. Within the new system, "Specified Response Questions" achieved higher stability ($M = 6.16$, $SD = 0.71$) compared to "Spontaneous Inquiries" ($M = 5.5$, $SD = 1.06$, $p = 0.0025$) , although both exceeded the quality benchmark of 5 ($p < 0.001$ and $p = 0.0083$, respectively). These findings confirm that integrating semantic retrieval with LLMs effectively bridges the semantic gap in audio data access, offering a stable and superior alternative to traditional keyword-based search methods.*

Keywords: Podcast Retrieval, Retrieval-Augmented Generation, Large Language Models (LLMs).

1. Introduction

While Large Language Models (LLMs) have become core assistive tools in daily life, they often suffer from "hallucinations" due to a lack of grounding in factual knowledge (Huang et al., 2025). To address this, Retrieval-Augmented Generation (RAG) has emerged as an effective framework that integrates semantic vector retrieval with generative models, significantly improving the accuracy and relevance of generated content (Liu & Cui, 2024).

Concurrently, podcasts have become a pivotal medium for information acquisition and systematic learning during fragmented time. However, traditional query systems rely on keyword matching within program titles or brief descriptions, failing to perform semantic-level searches on the audio content itself, which results in reduced learning efficiency.

To overcome these limitations, this study proposes an integrated RAG workflow: OpenAI Whisper is employed for high-precision speech-to-text transcription, and FAISS is utilized for efficient semantic retrieval by matching user queries with content vectors. Finally, the system leverages LLMs to analyze retrieved results and produce semantically aligned, contextually complete responses.

1.1. Research Objectives

This study investigates RAG integration to address semantic limitations in keyword-based podcast search, focusing on verifying technical feasibility within unstructured audio, statistically comparing retrieval efficiency against legacy models, and quantifying response reliability and user satisfaction across "Specified Response Questions" and "Spontaneous Inquiries".

2. Literature Review

Current Status and Limitations of Podcast Information Retrieval：Podcasts are primarily distributed via RSS feeds, which often lack standardized metadata and rely on informal naming conventions. Jones et al. (2021) noted that this lack of structure makes it difficult for users to locate specific content within episodes, especially since natural speech lacks explicit semantic classification. Consequently, RSS data is often considered "noisy" and unreliable. Furthermore, Tian et al. (2022) observed that platform search systems, such as Spotify's "Instant Search," struggle when users provide vague queries, forcing them to manually filter through voluminous results.

Evolution of Podcast Search Technology：Retrieval technology has transitioned from traditional text matching to semantic-level analysis. Early research by Mizuno et al. (2008) used confusion networks and TFp-IDFp weighting to mitigate speech recognition errors, achieving better accuracy than standard TF-IDF. Modern semantic retrieval utilizes

high-precision indexing; for instance, Krisnawati et al. (2024) demonstrated that IndexFlatL2 achieves maximum recall by performing full vector comparisons. Rusum et al. (2024) found that FAISS provides a high recall of 97.9% and balances query latency effectively, though it requires more storage to maintain precision.

Applications and Research Gap in RAG：Retrieval-Augmented Generation (RAG) enhances LLM outputs by providing external context, which reduces hallucinations.

However, He et al. (2025) pointed out that "Naive RAG" may still suffer from insufficient precision in specialized scenarios. Liu and Cui (2024) successfully implemented a document retrieval system using Word2Vec within a RAG framework to produce more rational answers. To capture complex relational knowledge, Zhou & Xiao (2025) utilized GraphRAG, though measuring the precision of its retrieved content remains challenging.

A significant research gap exists as traditional podcast systems rely on RSS-based keyword metadata, which fails to interpret complex natural language queries. Since mainstream platforms (e.g., Spotify, Apple Podcasts) share this similar keyword paradigm, Apple Podcasts is selected as a representative industry baseline to quantify the intergenerational performance gap between legacy indexing and frontier semantic RAG architectures. This study addresses this gap by integrating AI-driven semantic retrieval with RAG, enabling context-aware, human-like responses to complex natural language prompts.

3. Research Framework and Methodology

The research framework employs a three-phase workflow to transform unstructured podcast audio into a high-precision semantic search system: data preparation, system implementation, and performance evaluation. This process systematically converts raw audio into searchable text, implements retrieval-augmented generation via vector indexing, and concludes with a quantitative assessment of response quality and stability.

3.1. System Workflow Overview

The system implementation integrates several advanced technologies to ensure response accuracy:

- Data Processing: Podcast audio is downloaded via Python scripts and transcribed into text using the OpenAI Whisper "medium" model, which includes 60-second timestamps for precise location mapping. To ensure data quality, manual transcription correction is performed to fix proper noun errors.
- RAG Implementation: Transcripts are converted into 1024-dimensional vectors using the bge-m3 model. These are stored in a FAISS database using the IndexFlatL2 index to prioritize retrieval precision over speed.
- Response Generation: The llama-3.3-70b-versatile model generates responses in Traditional Chinese based on the retrieved context and user queries.

3.2. Evaluation Methodology

To assess efficacy, 30 participants rated responses on a 1–7 Likert scale, where 5 was predefined as the "Mostly Correct" quality benchmark. The evaluation involved two questionnaires: one comparing the New RAG vs. Legacy systems (10 items) and another assessing Specified vs. Spontaneous queries within the new system (10 items). This 20-item-per-participant structure yielded 600 total responses (30 participants across 20 items), providing a robust foundation for paired t-tests and one-tailed significance testing.

4. Results and Analysis

4.1. Descriptive Statistics

Statistical analysis was performed on the 600 item responses derived from the 30 questionnaires to evaluate performance. Preliminary analysis utilized descriptive metrics to reveal the central tendency and dispersion of user ratings.

4.1.1. Descriptive Statistics and Box Plot Analysis

Descriptive analysis reveals the central tendency and dispersion of user ratings on a 1–7 scale. The Legacy System averaged M=1.34 (SD=0.66), whereas the New System achieved M=6.06 (SD=0.86), demonstrating superior quality and consistency across diverse query scenarios. Within the new system, "Specified Response Questions" yielded higher stability (M=6.16, SD=0.71) compared to "Spontaneous Inquiries" (M=5.5, SD=1.06).

4.1.2. Inferential Statistics: Paired t-tests

Paired t-tests verified the statistical significance of observed differences. The New System significantly outperformed the legacy platform in response quality and stability ($t(29) = 20.018$, $p < 0.001$). Internal analysis confirmed that "Specified Response Questions" yielded significantly more stable and higher-quality responses than "Spontaneous Inquiries" ($t(29) = 3.307$, $p = 0.0025$).

4.2. One-tailed Test

One-tailed tests confirmed that average ratings significantly exceeded the benchmark of 5. As shown in Figure 1(a), "Specified Response Questions" yielded $t(29) = 8.744$ ($p < 0.001$), with the t-value falling within the rejection region. Similarly, "Spontaneous Inquiries"(Figure 1(b)) reached $t(29) = 2.545$ ($p = 0.0083$), demonstrating performance significantly above the threshold even in open-ended scenarios. These results collectively corroborate the system's stable, high-quality output across diverse query types.

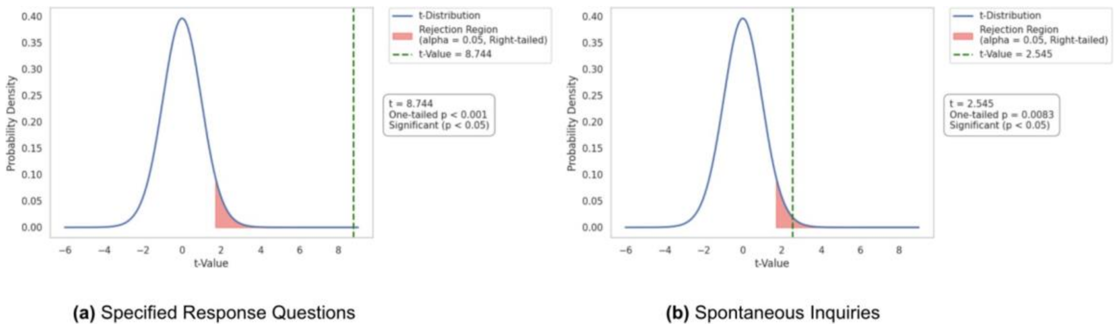


Figure 1. One-tailed significance tests for (a) Specified Response Questions and (b) Spontaneous Inquiries.

5. Conclusion

This study proposes a RAG system integrating OpenAI Whisper, bge-m3, FAISS, and Llama-3.3 to overcome the semantic limitations of traditional podcast search. Statistical analysis reveals that the New System (M=6.06, SD=0.86) significantly outperformed the Legacy platform (M=1.34, SD=0.66) with $t(29)=20.018$ and $p<0.001$. Internal comparisons confirmed that "Specified Response Questions" (M=6.16, SD=0.71) yielded more stable results than "Spontaneous Inquiries" (M=5.5, SD=1.06; $t(29)=3.307$, $p=0.0025$).

Furthermore, one-tailed tests verified that both question types significantly exceeded the quality benchmark of 5, with t-values of 8.744 ($p<0.001$) and 2.545 ($p=0.0083$), respectively. While transcription errors like homophones impact performance, future iterations utilizing LLM-based post-editing and embedding model optimization (e.g., text-embedding-3 vs.

bge-m3) will enhance scalability and retrieval efficiency. In conclusion, the RAG architecture effectively bridges the semantic gap in podcast search and offers a highly feasible direction for future audio data applications.

References

Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., ... & Liu, T. (2025). A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. ACM Transactions on Information Systems, 43(2), 1-55.

刘彦宏, & 崔永瑞. (2024). 基于 Word2Vec 模型与 RAG 框架的医疗检索增强生成算法. Artificial Intelligence and Robotics Research, 13, 479.

Jones, R., Zamani, H., Schedl, M., Chen, C. W., Reddy, S., Clifton, A., ... & Carterette, B. (2021, July).

Current challenges and future directions in podcast information access. In Proceedings of the 44th International acm sigir Conference on Research and Development in Information Retrieval (pp. 1554-1565).

Tian, M., Hauff, C., & Chandar, P. (2022, October). On the challenges of podcast search at spotify. In Proceedings of the 31st ACM International Conference on Information & Knowledge Management (pp. 5098-5099).

Mizuno, J., Ogata, J., & Goto, M. (2008, December). A similar content retrieval method for podcast episodes. In 2008 IEEE Spoken Language Technology Workshop (pp. 297-300). IEEE.

Krisnawati, L. D., Mahastama, A. W., Haw, S. C., Ng, K. W., & Naveen, P. (2024). Indonesian-english textual similarity detection using universal sentence encoder (USE) and facebook ai similarity search (FAISS). CommIT (Communication and Information Technology) Journal, 18(2), 183-195.

Rusum, G. P., & Anasuri, S. (2024). Vector Databases in Modern Applications: Real-Time Search, Recommendations, and Retrieval-Augmented Generation (RAG). International Journal of AI, BigData, Computational and Management Studies, 5(4), 124-136.

贺梓然, 江波, & 王晓龙. (2025). 改进检索增强与 LLM 思维链维修策略生成. 计算机应用与软件, 42(3), 1-6.

周俭航, & 肖诗斌. (2025). 图检索增强生成研究综述. Artificial Intelligence and Robotics Research, 14, 402.

Hsu, T. C., & Ogata, H. (Eds.). (2026). *Proceedings of the 10th International Conference on Artificial Intelligence, Computational Thinking and STEM Education* .Kumamoto: The Asia-Pacific Society for Computers in Education.

Influences of Generative AI Applications on Creative Self-Efficacy and Design Creativity

Ming-Yu LIN[1] & Yu-Shan CHANG[2*]
[1] Department of Industrial Education, National Taiwan Normal University, Taipei, Taiwan
[2] Department of Technology Application and Human Resource Development, National Taiwan Normal University, Taipei, Taiwan
*sam168@ntnu.edu.tw

Abstract: *The purpose of this study is to investigate the impact of Generative Artificial Intelligence (GenAI) applications on creative self-efficacy and design creativity. The study involved 32 first-year university students and employed a one-group pretest-posttest design to sequentially evaluate three different instructional methods. The findings indicate that GenAI has a positive impact on creative self-efficacy with a moderate effect size. This effect was primarily driven by improvements in "belief in creative products" and "resilience against negative evaluation," whereas no significant difference was found in "belief in creative thinking strategies." Furthermore, GenAI significantly enhanced design creativity—including technical performance, novelty, and aesthetic refinement—as well as overall creative output, all showing moderate to large effect sizes. Notably, integrating GenAI across all three stages of the process—material generation, ideation, and illustration generation—resulted in a positive impact with a large effect size.*

Keywords: Generative Artificial Intelligence (GenAI), Creative self-efficacy, design creativity

1. Introduction

According to a survey by the U.S. research firm PitchBook (2023), GenAI saw an explosion of investment and is now recognized as one of the most disruptive technologies of our time. This shift has profoundly impacted the educational sector, with recent research indicating that GenAI significantly enhances creative learning and critical thinking (Qian, 2025). Furthermore, a strong correlation exists between creative self-efficacy and overall creativity (Duran & Alejandro, 2025; Zhou & Lee, 2024); in the context of GenAI, this self-efficacy has emerged as a crucial factor influencing creative performance (Zhou et al., 2025). Consequently, the objective of this study is to investigate the impact of GenAI instructional applications on students' creative self-efficacy and their performance in design creativity.l.

2. Literature Review

GenAI demonstrates significant efficacy in creative education—specifically in creative stimulation, engagement, and output. However, the integration of GenAI also raises several critical concerns: (1)Over-reliance and Cognitive Offloading: Chronic dependence on GenAI may lead to intellectual passivity and reduced cognitive engagement, a phenomenon known as cognitive offloading. This may result in a growing aversion to deep thinking, ultimately undermining a student’s creative self-efficacy and innate creative abilities (Wei et al., 2025). (2)Reduction in Creative Diversity: While human-AI collaboration can elevate overall creative performance, it may paradoxically diminish the diversity of ideas. Users may become reliant on repetitive prompts and more inclined to passively accept AI-recommended concepts (Holzner et al., 2025). (3)Plagiarism and Assessment Challenges: GenAI introduces risks such as algorithmic bias, plagiarism, and "hallucinations" (incorrect answers). These issues make it increasingly difficult to assess individual contribution and authentic student engagement (Wall et al., 2025).

Design thinking is a methodical approach to problem-solving. Its framework typically consists of five core stages: empathize, define, ideate, prototype, and evaluate/refine (Harvard Business School, 2023; Hasso Plattner Institute of Design at Stanford, 2023), as illustrated in Figure 1 (Design Council, 2023; IDEO, 2023). The resulting creative outcomes—whether tangible products or strategic solutions—must be characterized by both novelty and appropriateness (Emami et al., 2023).

In other words, design creativity is essentially a continuous process of problem-solving (Sameti, Koslow, & Mashhady, 2022). Consequently, the assessment of creativity must account for both the process and the outcome (Esling & Devis, 2021). In the instructional design of this study, evaluations are conducted across the stages of definition and analysis, ideation, and final product output. These stages are assessed for innovation and aesthetic refinement to ensure the results align with the dual criteria of novelty and appropriateness (Yi et al., 2021).

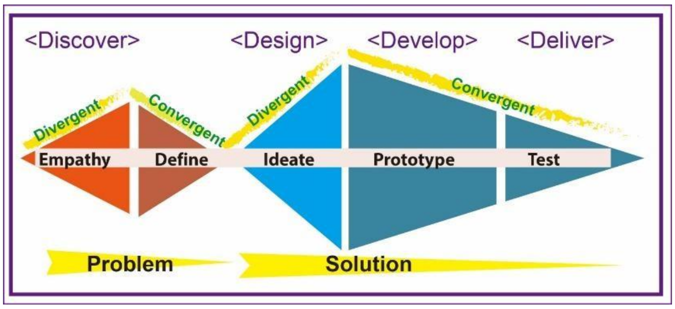


Figure 1. Framework of Design Thinking Process.
Note: Lin & Chang, 2024

3. Research design and implementation

3.1. Research participants: The participants in this study consisted of 32 freshmen from a prestigious public university in Taiwan.

3.2. Research process: The study began with a pre-test to assess students' creative self-efficacy. Subsequently, a unit on digital image creation was conducted using traditional instruction (TI). This was followed by the sequential implementation of two distinct teaching methods: GenAI-A and GenAI-B. At the conclusion of each instructional phase, students were required to produce an original digital image work.

3.3. Instruments: (1). Integrating GenAI into Design Creativity Instruction The instructional design followed the design thinking process, spanning the stages of definition and analysis, ideation, and final product output (IDEO, 2023). Creative outcomes were assessed based on novelty and appropriateness (Yi et al., 2021). Adobe Photoshop served as the primary software for teaching image processing and creation.

The three teaching methods are defined as follows: Traditional Instruction (TI): Students independently brainstormed concepts, sourced their own image assets, and executed the creative work manually. GenAI-A Method: Students independently developed their own creative concepts but utilized GenAI to generate primary image assets and backgrounds before completing the final composition manually. GenAI-B Method: Students engaged in a co-creative dialogue with GenAI to brainstorm concepts. GenAI was then used to generate primary image assets, backgrounds, and reference illustrations. Finally, students performed creative optimization and manual refinement to produce the final work.

Design Creativity Rubric Based on the established criteria of effectiveness, novelty, and aesthetic refinement in design (Besemer, 1998; Mazerant et al., 2021; Weisberg et al., 2021), this study developed a scoring rubric comprising three items: Skills & Elaboration, Novelty, and Aesthetics. The overall scoring reliability reached a Cronbach's alpha of .813, indicating robust reliability and validity (Hair et al., 2010).

		
Figure 2. Sample work from the TI (Traditional Instruction) group. Description: A skull covered in screen notifications, symbolizing the omnipresence of digital information.	*Figure 3*. Sample work from the GenAI-A group. Description: "I lit an eternal lamp for you, hoping my longing reaches you from afar. I vaguely remember you saying the moon is fullest on the fifteenth... now, whenever I miss you, I simply look up at the sky."	*Figure 4*. Sample work from the GenAI-B group. Description: This composition utilizes an X-ray transparency effect overlaid with images of electronic components to create a sense of depth. It illustrates the fusion of mechanical and natural structures—a concept piece for a cyber-insect.

Creative Self-Efficacy Scale This study utilized the Creative Self-Efficacy Scale developed by Lin & Chang (2025). The scale consists of 12 items across three dimensions: Creative Thinking Strategy Beliefs (e.g., "When brainstorming, I believe I can generate a wide variety of ideas"), Creative Product Beliefs (e.g., "I believe the products I design are more creative than similar ones on the market"), and Resilience Against Negative Evaluation (e.g., "I will continue to think outside the box even if my unique ideas are not initially accepted"). According to a Principal Component Analysis (PCA), the factor loadings (and percentage of variance accounted for) for the three components were 6.69 (47.81%), 1.06 (7.61%), and 1.01 (7.24%), respectively. The total explained variance reached 62.67%, which is considered acceptable (UCLA, 2021). The reliability coefficients (Cronbach's alpha) for the three dimensions were .758, .838, and .656, with an overall reliability of .895 (Lin & Chang, 2025), demonstrating strong reliability and validity (Hair et al., 2010).

3.4. Data Analysis: The primary statistical method employed in this study was the paired samples t-test. This analysis was used to determine the impact of GenAI instructional applications on students' design creativity.

3.5. Research Ethics: This study adhered to the ethical principles and regulations outlined in the Declaration of Helsinki.

4. Results

4.1. The Impact of GenAI on Creative Self-Efficacy : A paired samples t-test revealed that integrating GenAI into design creativity instruction led to a significant increase in creative self-efficacy ($t = 2.308$, $p = .014$), with a moderate effect size (Cohen's $d = .426$). This primary effect was driven by improvements in creative product beliefs and resilience against negative evaluation, whereas no significant difference was observed in creative thinking strategy beliefs.

4.2. The Impact of GenAI on Design Creativity Performance: A paired samples t-test revealed the following:GenAI-A vs. TI: Compared to the Traditional Instruction (TI) method, the GenAI-A method led to significantly higher creativity performance ($t = 2.576$, $p = .009$), with a moderate effect size (Cohen's $d = .549$). Moderate positive effects were observed specifically in technical performance, novelty, and aesthetic refinement.GenAI-B vs. TI: The GenAI-B method also outperformed the TI method ($t = 7.831$, $p < .001$) with a large effect size (Cohen's $d = 1.670$). While technical performance and aesthetic refinement showed moderate positive effects, novelty demonstrated a large effect size (Cohen's $d = 1.005$).GenAI-B vs. GenAI-A: Comparing the two AI-assisted methods, there were no significant differences in the individual dimensions of technical performance, novelty, or aesthetic refinement. However, the aggregate total score showed that GenAI-B was significantly superior to GenAI-A ($t = 3.945$, $p < .001$).

5. Conclusions

5.1. GenAI’s Moderate Positive Impact on Creative Self-Efficacy : The findings of this study demonstrate that Generative AI (GenAI) has a positive impact on creative self-efficacy with a moderate effect size. This effect is primarily attributed to improvements in creative product beliefs and resilience against negative evaluation, whereas creative thinking strategy beliefs showed no significant difference. These results align with research suggesting that GenAI fosters creative stimulation and supports every stage of the production process (Newman et al., 2024), often serving as a key mediator for creative performance (Zhou et al., 2025). However, it is important to heed warnings that learners may perceive GenAI merely as a "ready-made output machine." Viewing AI as a utility tool rather than a creative partner may inadvertently stifle a student's own imaginative capacity (Newman et al., 2024).

5.2. GenAI’s Significant Impact on Design Creativity Performance: This study found that GenAI applications exert a moderate-to-large positive effect on technical performance, novelty, and aesthetic refinement, as well as overall design creativity. Notably, the GenAI-B method achieved a large effect size in both novelty and total performance. These findings reinforce the growing consensus that AI integration enhances overall creativity and creative output (Duran & Alejandro, 2025; Zhou & Lee, 2024).

Interestingly, while there were no significant differences in the individual dimensions (technical performance, novelty, and aesthetics) between the GenAI-A method (limited to asset generation) and the GenAI-B method (spanning asset generation, ideation, and illustration generation), a significant difference did emerge in the aggregate total scores. This suggests that while GenAI is undeniably beneficial for design creativity, future research should utilize multivariate analysis (MANOVA) to provide a more holistic and in-depth evaluation of design creativity as a multi-dimensional construct.

6. Limitations and Suggestions

Through experimental teaching, this study demonstrates that the application of GenAI in instruction yields significant positive effects on design creativity. Based on these findings, it is recommended that universities increase the integration of GenAI across various curricula. Furthermore, in light of the study's constraints, the following suggestions are proposed for future research:

1. Limitations in Research Design: This study employed a one-group pretest-posttest design, which is categorized as a pre-experimental design, and utilized paired samples t-tests for data analysis. This design is inherently susceptible to maturation effects and testing effects (practice effects). To achieve more precise experimental results, future studies should incorporate a control group to better isolate the impact of the intervention.
2. Limitations in Analytical Methods: This study used t-tests to compare individual sub-scores and total scores of creative performance. This approach may be affected by the cumulative effect or sub-dimension summation effect, potentially impacting statistical power and measurement error. Future research is encouraged to utilize multivariate analysis (e.g., MANOVA) to enhance discriminatory power and provide a more robust evaluation of the relationships between variables.

References

Besemer, S. P. (1998). Creative product analysis mat rix: Testing the model structure and a comparison among products three novel chairs. *Creativity Research Journal, 11*(4), 333 346.

(Additional references are available from the author upon request.)

Classroom Practices of Intelligent Affective Education: A Study of AI Integration in SEL Curriculum Models Based on Bibliometric and Content Analysis

Hsiang-Ni Tsai[1],Sheng-Yi Wu[2]
[1,2] National Tsing Hua University, Hsinchu, Taiwan
iamwiifit1@gmail.com, digschool@gmail.com

Abstract: *This study explores the integration of Generative AI (GenAI) into Social and Emotional Learning (SEL) to address rising mental health needs. Using bibliometric analysis (n=122) and case studies (n=15), the research identifies a shift from technical monitoring to deep disciplinary integration in fields like STEM and Arts. Findings reveal that AI serves as a simulator, reflection guide, and diagnostician, significantly enhancing students' self-awareness and decision-making. The study concludes that successful humanistic intelligent education requires strengthening teachers' TPACK and establishing robust ethical frameworks for human-AI collaboration.*

Keywords: Generative AI, Social and Emotional Learning (SEL), TPACK, Affective Education, Literature Analysis

1. Introduction

In the post-pandemic era, students' needs for mental health and social-emotional learning (SEL) have shown exponential growth (Tyler et al., 2025). However, the technological convenience brought by the digital shift has also posed new challenges to the quality of interpersonal interaction among students, while simultaneously leading to a loss of the ability for deep thinking on subjects and issues (Arinushkina et al., 2023; Banda & Czako, 2026). Although Taiwanese students perform excellently in international academic assessments (Lin et al., 2021), they remain relatively weak in affective dimensions such as learning motivation and self-confidence (OECD, 2016). To this end, the Ministry of Education in Taiwan launched the "Social and Emotional Learning Medium-to-Long-Term Plan (2025–2029)" in 2025(Ministry of Education, 2025), aiming to incorporate SEL into the current curriculum guidelines. However, the practical field faces challenges such as uneven resource allocation, differences in teachers' affective teaching literacy, and the difficulty of providing real-time, individualized guidance (Ministry of Education, Taiwan, 2025).

The rise of Generative AI has provided a safe, low-risk, and high-frequency interaction practice environment for SEL, which can effectively compensate for the limitations of traditional teaching in social simulation (Lee & Lee, 2025). Although GenAI has shown effectiveness in cognitive learning assistance, its systematic integration models, pedagogical role positioning (such as guider, digital peer, or evaluator), and specific integration paths in affective education remain to be clarified. This study first adopts bibliometrics to map the development landscape of AI and SEL curriculum integration between 2020 and 2026; subsequently, it utilizes content analysis to deeply explore the current implementation status and future trends of AI-SEL integrated teaching models.

2. Literature Review

Social and emotional learning has received significant attention and emphasis in recent years, transforming from supplementary counseling content into the core of the educational system; its core components include self-awareness, self-management, social awareness, relationship skills, and responsible decision-making (CASEL, 2019). However, as digital natives who participated heavily in digital learning during the pandemic present a highly online lifestyle post-pandemic, student mental health faces challenges such as high levels of attention fragmentation, social isolation, and declining psychological resilience (Buzzi et al., 2025). This makes the promotion of the "Social and Emotional Learning Medium-to-Long-Term Plan (2025–2029)" by the Ministry of Education in Taiwan particularly important. Currently, research has pointed out that the promotion of affective education should no longer be decoupled from disciplinary knowledge but should be integrated with various subjects, establishing an intelligent learning environment for students' emotional regulation through curriculum arrangements (Tang & Herrera, 2025).

With the rise of Generative AI, past research on artificial intelligence and emotion mostly focused on external detection and monitoring, where researchers used AI to capture detailed facial changes or emotional shifts; such research trends emphasized stress prediction based on physiological data (Bilucaglia et al., 2024). However, current artificial intelligence can achieve large amounts of dialogue, responses, and even empathy, and can also perform role-playing to guide students toward a deeper sense of immersion in various issues. This mechanism also helps students practice moral judgment and emotional regulation in a relatively low-risk environment (Aprilia et al., 2025). Additionally, in the extended application of the TPACK framework, AI not only optimizes pedagogical knowledge (PK) but also changes current teaching methods (TPK), precisely integrating technology and affect into current teaching (Cohn et al., 2024). In the past, technology was seen as a representative of rationality, efficiency, and scientification, while concepts such as emotion and affect were abstract and perceptual; the fusion of the two in the educational field will bring a brand-new face to future disciplinary content.

3. Research Methodology

This study explores through literature analysis methods, divided into two major steps. First, a bibliometric analysis was conducted. In the Scopus database, the conditions were set into three major parts: technical means (AI), core targets (SEL), and teaching practice. The technical means part included keywords such as AI, Generative AI, Large Language Models (LLM), Chatbots, and Affective Computing; core target keywords included Social Emotional Learning, Emotional Intelligence (EQ), and Well-being; education practice keywords included Curriculum, Instructional Design, Pedagogy, Teaching Models, and Classroom Integration. The timeframe was set from 2020 to 2026. Since 2022 was the key year when Generative AI emerged and began to flourish, this study set the analysis for the years before and after the emergence of Generative AI. Simultaneously, English-language journals or conference papers were used as the screening mechanism, resulting in 122 articles. In the second part, this study screened 15 articles from the 122 for further content analysis to explore the specific teaching models of current AI-SEL curriculum fusion.

4. Results and Discussion

4.1. Global Trends in AI-Integrated Affective Education

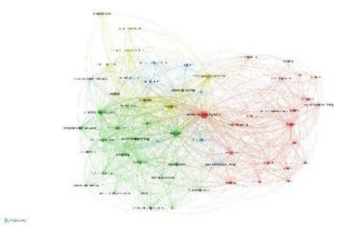


Figure1. Network Diagram

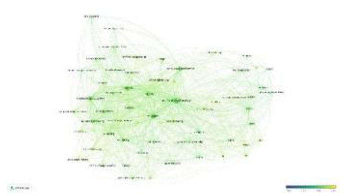


Figure2 Time series diagram

This study explores through literature analysis methods, divided into two major steps. First, a bibliometric analysis was conducted. In the Scopus database, the conditions were set into three major parts: technical means (AI), core targets

(SEL), and teaching practice. The technical means part included keywords such as AI, Generative AI, Large Language Models (LLM), Chatbots, and Affective Computing; core target keywords included Social Emotional Learning, Emotional Intelligence (EQ), and Well-being; education practice keywords included Curriculum, Instructional Design, Pedagogy, Teaching Models, and Classroom Integration. The timeframe was set from 2020 to 2026. Since 2022 was the key year when Generative AI emerged and began to flourish, this study set the analysis for the years before and after the emergence of Generative AI. Simultaneously, English-language journals or conference papers were used as the screening mechanism, resulting in 122 articles. In the second part, this study screened 15 articles from the 122 for further content analysis to explore the specific teaching models of current AI-SEL curriculum fusion.

4.2. Teaching Integration Fields and Pedagogical Roles

In the distribution of teaching integration fields, the 15 core documents analyzed in this study show a clear trend of disciplinary integration. Among them, 11 studies belong to the disciplinary integration model, applying AI technology widely to professional fields such as architectural design, art education, and STEM+C (Zahra et al., 2025; Aprilia et al., 2025; Cohn et al., 2024). This data confirms that affective literacy is not isolated knowledge transmission but, through AI as a technical mediator, allows students to achieve emotional regulation and social literacy goals while acquiring professional subject knowledge. In addition, the research fields also extend to specialized courses centered on "life shaping" and "health and well-being" (2 articles) (Zadorozhna-Kniahnytska et al., 2025; Dekker et al., 2020), as well as counseling mechanisms providing individualized emotional support outside classroom hours (3 articles) (Sturgill et al., 2021; Tariq et al., 2025; Dekker et al., 2020). Such diverse field distribution illustrates that intelligent affective education is shifting from single counseling room interventions to all-around curriculum penetration, developing a teaching status with contextualized characteristics.

Regarding the pedagogical roles played by AI in the curriculum, this study summarizes three main practice scenarios. First is the situation simulator and practitioner; researchers utilize the role-playing characteristics of Generative AI to let students dialogue with virtual characters (such as fishermen or scientists) to practice cross-perspective thinking (Nguyen et al., 2025), or construct moral dilemma scenarios for decision-making drills, achieving low-risk affective practice (Amini et al., 2024). Second is the reflection guider; a total of 4 documents emphasize the heuristic function of AI, guiding students to perform mindfulness breathing, optimize writing arguments, or clarify research ideas, thereby strengthening self-awareness (Sturgill et al., 2021; Zhang, 2025; Aure & Cuenca, 2024; Dekker et al., 2020). Finally, the highest proportion is the monitor and diagnostician; a total of 5 studies utilize multimodal data technology to analyze students' concentration and stress levels through video, voice, and physiological data, instantly detecting their psychological load and emotional needs during collaborative or design tasks (Zahra et al., 2025; Zhang & Leong, 2025; Tariq et al., 2025; Aprilia et al., 2025; Cohn et al., 2024). These three types of roles demonstrate how AI has progressively advanced from external emotional monitoring to internal emotional guidance, becoming an indispensable intelligent coach in the curriculum.

4.3. Pedagogical Models and SEL Core Literacy

In the dimension of integration between pedagogical methods and content knowledge (TPACK), the research presents diverse practice paths. First, in terms of gamified guidance (TPK1), electronic textbooks significantly enhance students' motivation to participate in safety education courses by integrating multimedia and interactive tasks (Zadorozhna-Kniahnytska et al., 2025). Second, in the application of scaffolded reflection (TPK2), a total of 4 documents emphasize the importance of metacognition, pointing out that students need to possess the awareness to evaluate and integrate resources when using Generative AI for creation, and deepen self-reflection through writing journals guided by LLMs (Karthikeyan et al., 2025; Zhang, 2025; Aure & Cuenca, 2024; Dekker et al., 2020). Furthermore, the 5 studies on adaptive paths (TPK3) demonstrate how systems recommend mindfulness courses based on users' anxiety levels, or instantly trigger teacher intervention mechanisms when detecting a drop in student concentration or encounters with difficulties, embodying technical support for teaching flexibility (Sturgill et al., 2021; Zahra et al., 2025; Zhang & Leong, 2025; Aprilia et al., 2025; Cohn et al., 2024). Finally, in the part of affective content transformation (TCK1), researchers further translate values such as identity and sense of connectedness into design principles for chatbots, or integrate moral emotion theories into English teaching, achieving deep alignment between disciplinary content and affective values (Nguyen et al., 2025; Amini et al., 2024).

Regarding the implementation of Social and Emotional Learning (SEL) core literacy, research results mainly focus on three major dimensions. In self-awareness and management, 6 documents confirm that AI intervention can effectively reduce students' anxiety and depression, and through the regulation of emotional intelligence and self-esteem, improve their perceived psychological health, academic confidence, and subjective well-being, strengthening individuals' stress management and self-regulation abilities (Sturgill et al., 2021; Zhao & Liu, 2025; Zhang, 2025; Aure & Cuenca, 2024; Tariq et al., 2025; Dekker et al., 2020). In social awareness and relationship skills, 5 studies focus on emotional cues in teacher-student interactions, using AI to establish a sense of connection between students and different roles, thereby enhancing empathy and analyzing interaction patterns in collaborative learning to ensure fairness of participation (Zahra et al., 2025; Nguyen et al., 2025; Aprilia et al., 2025; Finkelstein & Soffer-Vital, 2025; Cohn et al., 2024). As for responsible decision-making, 3 articles deeply explore dependency and academic integrity risks in AI use, emphasizing that students should develop a sense of responsibility for environmental sustainability, moral judgment, and correct attitudes toward technology use in Generative AI research (Kniahnytska et al., 2025; Nguyen et al., 2025; Amini et al., 2024).

4.4. Fusion Challenges and Ethics

In exploring the challenges and ethics of the fusion process, the research reveals dual gaps in technical and pedagogical aspects. First, at the level of technical and psychological risks, two articles warn that students' excessive reliance on AI may lead to loss of creativity and cognitive laziness, and emphasize the importance of developing "Explainable Artificial Intelligence (XAI)" to avoid black-box decisions and eliminate anxiety and misunderstanding caused by the lack of non-verbal cues in digital environments (Kniahnytska et al., 2025; Tariq et al., 2025; Finkelstein & Soffer-Vital, 2025). Furthermore, the TPACK literacy gap of teachers is also a key challenge; three studies consistently point out that teachers need to possess a composite ability to integrate psychological literacy, technical knowledge, and intelligent ethics, and must deeply participate in the human-machine collaboration process to accurately interpret emotional data produced by AI and subsequently provide feedback and counseling with warmth (Zhao & Liu, 2025; Zhang & Leong, 2025; Cohn et al., 2024).

5. Conclusions and Recommendation

Smart affective education is undergoing a paradigm shift from technical monitoring toward systemic curricular integration. Following the explosion of Generative AI, the research focus has transitioned from algorithmic decision-making to disciplinary infiltration, positioning AI as an instructional mediator and an "implicit curriculum" (Kniahnytska et al., 2025; Zhao & Liu, 2025). This evolution precisely addresses the demands for individualized and real-time guidance outlined in the 2025–2029 SEL Strategic Plan. Within the TPACK framework, AI provides diverse pedagogical pathways through gamification (Nguyen et al., 2025) and scaffolded reflection (Sturgill et al., 2021). It has evolved from an external monitoring tool into an affective support system capable of actively triggering student self-awareness and regulation (Cohn et al., 2024), thereby strengthening self-management, social skills, and a sense of connection in collaborative learning.

While AI shows significant potential in enhancing psychological well-being, risks such as the loss of creativity due to over-reliance and the absence of non-verbal cues in digital environments must be mitigated through the transparent decision-making mechanisms of Explainable AI (XAI) (Tariq et al., 2024). Ultimately, the success of smart affective education hinges on the pivotal role of the teacher. Educators must possess a composite competency that integrates psychological literacy with technical knowledge to interpret affective data with human "warmth." Consequently, future development should focus on establishing ethical frameworks for human-AI collaboration and enhancing teachers' affective guidance efficacy within digital transformations (Zhang & Leong, 2025), ensuring the integrity of technology-enhanced moral judgment and well-being.

References

Amini, M., Zahabi, A., & Wang, Y. (2024). *Employing artificial intelligence to foster moral emotions in English language education: A review*. In V. P. H. Pham, A. Lian, A. Lian, & J. White (Eds.), Empowering educators: Integrating AI tools for personalized language instruction (Vol. 1228, pp. 109–124). Springer. https://doi.org/10.1007/978-3-032-01348-4_2

Aprilia, P., Ijudin, M. F., Manda, A., & Adela, D. (2025). Utilization of emotional artificial intelligence (EAI) in art learning to improve character education of elementary school students. *Engineering Proceedings*, *107*(1), 135. https://doi.org/10.3390/engproc2025107135

Arinushkina, A. A., Morozov, A. V., & Robert, I. V. (Eds.). (2023). *The impact of digitalization in a changing educational environment*. IGI Global. http://doi.org/10.4018/979-8-3693-0433-4

Aure, P. A., & Cuenca, O. (2024). Fostering social-emotional learning through human-centered use of generative AI in business research education: An insider case study. *Journal of Research in Innovative Teaching & Learning*, *17*(2), 168–184. https://doi.org/10.1108/JRIT-03-2024-0076

Banda, P., & Czako, K. (2026). *The impact of digital transformation on interpersonal relationships and mental well-being in higher education institutions*. In: El Khoury, R. (eds) Business Models of the Future. Studies in Systems, Decision and Control, vol 238. Springer, Cham. https://doi.org/10.1007/978-3-031-85398-2_83

Bilucaglia, M., Casiraghi, C., Bruno, A., Chiarelli, S., Fici, A., Russo, V., & Zito, M. (2024, September). *Emotional reactions to AI-generated images: A pilot study using neurophysiological measures*. In: Nicosia, G., Ojha, V., Giesselbach, S., Pardalos, M.P., Umeton, R. (eds) Machine Learning, Optimization, and Data Science. LOD ACAIN 2024 2024. Lecture Notes in Computer Science, vol 15510. Springer, Cham. https://doi.org/10.1007/978-3-031-82487-6_11

Buzzi, M. C., Buzzi, M., Leporini, B., Bastiani, L., Denoth, F., Franchini, M., Pieroni, S., & Molinaro, S. (2025, January). *The impact of technology on Italian students during the COVID-19 pandemic: Learning, emotions, behaviour and support.* In Proceedings of the 21st International Conference on Web Information Systems and Technologies. SciTePress. https://doi.org/10.5220/0013778900003985

CASEL. (2019). *CASEL's framework for systemic social and emotional learning*. Frameworks - AWG - Measuring SEL. https://measuringsel.casel.org/wp-content/uploads/2019/08/AWG-Framework-Series-B.2.pdf

Cohn, C., Snyder, C., Fonteles, J. H., Ashwin, T. S., Montenegro, J., & Biswas, G. (2024). A multimodal approach to support teacher, researcher and AI collaboration in STEM+C learning environments. *British Journal of Educational Technology*. Advance online publication. https://doi.org/10.1111/bjet.13518

Dekker, I., De Jong, E. M., Schippers, M. C., De Bruijn-Smolders, M., Alexiou, A., & Giesbers, B. (2020). Optimizing students' mental health and academic performance: AI-enhanced life crafting. *Frontiers in Psychology*, *11*, Article 1063. https://doi.org/10.3389/fpsyg.2020.01063

Finkelstein, I., & Soffer-Vital, S. (2025). Cultural empathy in AI-supported collaborative learning: Advancing inclusive digital learning in higher education. *Education Sciences*, *15*(10), 1305. https://doi.org/10.3390/educsci15101305

Karthikeyan, K., Vasudevan, A., Sowmya, K. R., Jamil, N. A., & Hanefar, S. B. M. (2025). AI-enabled creativity: Effects on divergent and convergent thinking among higher education students. *International Journal of Engineering Pedagogy (iJEP)*, *15*(7), 68–85. https://doi.org/10.3991/ijep.v15i7.59051

Lee, Y., & Lee, J. (2025). Investigating conversational patterns with generative AI NPCs in role-play for elementary students' social and emotional learning. *CEUR Workshop Proceedings*, *3995*, 251–

256. https://ceur-ws.org/Vol-3995/GBLA25_paper2.pdf

Lin, S. W., Tzou, H. I., Lu, I. C., & Hung, P. H. (2021). Taiwan: Performance in the programme for international student assessment. *Improving a Country's Education*, 203. https://link.springer.com/chapter/10.1007/978-3-030-59031-4_10

Ministry of Education, Taiwan. 2025. The First Phase (114 — 118) of the MOE' sMedium- and Long-Term Plan for Social Emotional Learning. Retrieved August 29, 2025 from https://cfte.site.nthu.edu.tw/var/file/275/1275/img/127742991.pdf

Naixin, Z., & Leong, W. Y. (2025). Integrating artificial intelligence into pedagogy: Theoretical framework

and application model of I-TPACK in intelligent education. In W. Y. Leong (Ed.), *Industry 5.0: Design, standards, techniques and applications for manufacturing* (Chapter 17). Institution of Engineering and Technology.https://doi.org/10.1049/PBME026E_ch17

Nguyen, H., Nguyen, V., Ludovise, S., & Santagata, R. (2025). Value-sensitive design of chatbots in environmental education: Supporting identity, connectedness, well-being and sustainability. *British Journal of Educational Technology*. Advance online publication. https://doi.org/10.1111/bjet.13568 OECD. (2016). *PISA 2015 results: Excellence and equity in education* (Vol. I). PISA, OECD Publishing.

https://doi.org/10.1787/9789264266490-en

Sturgill, R., Martinasek, M., Schmidt, T., & Goyal, R. (2021). A novel artificial intelligence-powered emotional intelligence and mindfulness app (Ajivar) for the college student population during the COVID-19 pandemic: Quantitative questionnaire study. *JMIR Formative Research*, *5*(1), Article e25372. https://doi.org/10.2196/25372

Tang, A., & Herrera, L. J. P. (2025). Integrating affective knowledge into TPACK: A new visual representation. *Technology in Language Teaching & Learning*, *7*(2), 102852. https://doi.org/10.29140/tltl.v7n2.102852

Tariq, R., Orozco-del-Castillo, M. G., Zamir, M. T., Ramírez-Montoya, M. S., & Wilberforce, T. (2025). Explainable artificial intelligence for predictive modeling of student stress in higher education. *Scientific Reports*, *15*, Article 38375. https://doi.org/10.1038/s41598-025-22171-3

Tyler, P. M., Hurley, K. D., Mitchell, A. K., Chmelka, M. B., & Bharwani, S. W. (2025). Educator perspectives on safe and effective classrooms for students. *Education and Treatment of Children*, 1–17. https://doi.org/10.1007/s43494-025-00156-1

Zadorozhna-Kniahnytska, L., Goliuk, O., Hroshovenko, O., Prysiazhniuk, L., & Khadzhynova, I. (2025). *Usage of an interactive electronic textbook application: Reflection of Ukrainian experience (Based on the integrated course "Health, Safety, and Well-Being").* In Environment. Technology. Resources. Proceedings of the 16th International Scientific and Practical Conference (Vol. 3, pp. 375–381). http://doi.org/10.17770/etr2025vol3.8528

Zahra, S., Samra, M., & El Gizawi, L. (2025). Working toward advanced architectural education: Developing an AI-based model to improve emotional intelligence in education. *Buildings*, *15*(2),

356.https://research.ebsco.com/c/saa2ze/search/details/snvsigkk5z?request-context=plink&db=asi

Zhang, M. (2025). Optimizing academic engagement and mental health through AI: An experimental study on LLM integration in higher education. *Frontiers in Psychology*, *16*, Article 1641212. https://doi.org/10.3389/fpsyg.2025.1641212

Zhao, J., Li, X., Wei, J., Long, X., & Gao, Z. (2025). Understanding the psychological pathways to translation technology competence: Emotional intelligence, self-esteem, and innovation capability among EFL students. *BMC Psychology*, *13*, Article 66. https://doi.org/10.1186/s40359-025-02400-0

Hsu, T. C., & Ogata, H. (Eds.). (2026). *Proceedings of the 10th International Conference on Artificial Intelligence, Computational Thinking and STEM Education* .Kumamoto: The Asia-Pacific Society for Computers in Education.

Mapping Thematic Structures of Secondary School Invention Education Research in Korea (2020–2025): A Topic Modeling Analysis

Nanhee RYU[1] & Jeongmin LEE[2*]
[1,2] Department of Educational Technology, Ewha Womans University, Republic of Korea
*jeongmin@ewha.ac.kr

Abstract: *This study examines recent thematic trends in secondary school invention education research in Korea during the six-year period from 2020 to 2025. In recent years, invention education has increasingly intersected with artificial intelligence, maker education, and entrepreneurship as part of broader educational innovation. To identify dominant thematic structures, 80 KCI-indexed journal articles were analyzed using Latent Dirichlet Allocation (LDA) topic modeling. Five major topic clusters were identified: (1) AI- and maker-based problem-solving program development, (2) teaching-learning design, teachers, and policy support for invention education, (3) entrepreneurship and startup competency education, (4) curriculum and implementation of invention and intellectual property education, and (5) educational environments, infrastructure, and tools for invention education. The findings reveal a strong emphasis on program development and instructional innovation in secondary school invention education, while longitudinal and institutional integration research remains comparatively limited. These results suggest the need for stronger curriculum integration, sustained teacher and policy support, and broader implementation frameworks for secondary school invention education.*

Keywords: invention education, secondary school, research trends, topic modeling

1. Introduction

In an era shaped by rapid technological change, including artificial intelligence, robotics, and digital transformation, students increasingly need opportunities to engage with science and technology in meaningful ways, identify real-world problems, and develop practical solutions. Invention education serves as an educational approach that enables such learning experiences (Lee & Tae, 2013). A previous study has proposed a continuous framework for invention education across elementary, middle, and high school levels and emphasized the need for systematic education tailored to each school stage (Scharon et al., 2024). However, research on invention education in Korea has not been conducted in a balanced manner across different school levels (Lee, 2015).

Secondary school is a particularly important period during which students begin to consolidate their academic interests, career aspirations, and self-directed learning capacities. Prior research suggests that career-linked invention education can positively influence creative thinking and career development competencies (Kim & Park, 2025). Nevertheless, research at the secondary level remains relatively limited due to examination-oriented educational structures and constrained operational environments (Dalela & Ahmed, 2024; Moon, 2022).

Meanwhile, significant changes have occurred in the environment of invention education since 2020. The COVID-19 pandemic accelerated the transition from face-to-face programs to online and hybrid formats (Kim, 2022; Michael et al., 2021). Additionally, invention education has increasingly intersected with digital tools, artificial intelligence, maker education, intellectual property education, and entrepreneurship initiatives following national policy expansion (Korean Intellectual Property Office, 2022). Despite these structural and technological shifts, systematic, data-driven mapping of recent secondary school invention education research in Korea remains limited.

Accordingly, this study applies topic modeling to KCI-indexed journal articles published between 2020 and 2025 to examine the thematic structures of secondary school invention education research in Korea. The research question is as follows: What major topics emerge from topic modeling of secondary school invention education research in Korea from 2020 to 2025, and what do these topics suggest about the current direction and future development of the field?

2. Background

Invention education refers to educational efforts aimed at developing creativity, problem-solving ability, critical thinking, and intellectual property literacy (Korean Intellectual Property Office, 2017). It seeks to cultivate creative and convergent talent capable of responding to rapid technological and social change (Korean Intellectual Property Office, 2022).
In this study, invention education is defined as an educational process grounded in inventive thinking, in which learners identify problems, generate ideas, design and implement solutions, and refine outcomes based on creative and practical reasoning (Canedo, 1989). This conceptualization situates invention education as an integrative domain connecting maker education, STEAM education, intellectual property education, and entrepreneurship education through a shared emphasis on creative problem solving and value creation.

3. Research Methods

3.1. Research Procedure

This study analyzed domestic academic articles published between 2020 and 2025 to examine recent research patterns in secondary school invention education in Korea. For this purpose, the Research Information Sharing Service (RISS) was used as the main database, and only articles published in KCI-indexed journals were retained during the screening process. The search keywords included “invention education,” “maker education,” “creative convergence education,” “STEAM education,” “intellectual property education,” and “entrepreneurship education,” and the publication period was limited to 2020–2025. As a result of this search, a total of 710 articles were initially identified (search completed as of March 2026). Based on predetermined inclusion and exclusion criteria, a two-stage screening process focusing on titles and abstracts was conducted, followed by a full-text review. Through this process, 80 articles were included in the final analysis.

Table 1. *Inclusion and Exclusion Criteria*

Inclusion Criteria	Exclusion Criteria
Published in KCI between 2020 and 2025	Published before 2020 or after 2025
Written in Korean	Written in other language
Targeting secondary education	Not targeting secondary education
Invention education studies	Studies not related to education
Studies within maker, creative convergence, STEAM, intellectual property, or entrepreneurship education that include elements of inventive thinking	Studies focusing solely on maker, creative convergence, STEAM, intellectual property, or entrepreneurship education without elements of inventive thinking

3.2. Topic Modeling Methods

Topic modeling was conducted in R. The titles and abstracts of the selected articles were collected and preprocessed through noun-based morphological analysis and stopword removal using the KoNLP package. A document-term matrix (DTM) was constructed, and the Latent Dirichlet Allocation (LDA) algorithm was applied. The optimal number of topics was examined using the ldatuning package by comparing candidate models with k values from 2 to 5. Based on the Deveaud2014 metric, the model with five topics (k = 5) showed the highest coherence and was selected as the final model.

Although the corpus size was relatively small (n = 80), LDA was employed as an exploratory mapping tool to identify thematic structures rather than to generate predictive generalizations.

4. Results

This study conducted LDA-based topic modeling using the titles and abstracts of 80 domestic academic articles related to secondary school invention education research and classified them into five topics. The distribution of documents across the five topics indicated that Topic 1 accounted for the largest proportion of studies, followed by Topic 3, Topic 4, Topic 5 and Topic 2, suggesting a stronger concentration on program-level development research within the field.

The top keywords for Topic 1 included “development”, “problem”, “students”, “technology”, “maker”, “artificial intelligence”, and “STEAM”. Based on these keywords, the topic was labeled “AI- and maker-based problem-solving program development”. The co-occurrence of terms related to program development, students, and maker or AI suggests that this topic reflects a body of research focused on developing educational programs that incorporate maker activities and artificial intelligence elements.

The top keywords for Topic 2 included “invention education,” “teachers,” “stages,” “teaching,” “design,” and “policies.” Based on these keywords, the topic was labeled “teaching-learning design, teachers, and policy support for invention education.” Although this topic accounted for the smallest proportion of the corpus, it indicates a line of research addressing the role of teachers and policy support in invention education.

The top keywords for Topic 3 included “entrepreneurship”, “competency”, “learning”, “class”, “model”, “career”, and “startup”. Based on these keywords, the topic was identified as “entrepreneurship and startup competency education”. This indicates that secondary invention education research has recently expanded beyond simple idea creation and production activities to entrepreneurship and startup education.

The top keywords for Topic 4 included “invention”, “knowledge”, “development”, “property”, “subjects”, “high school”, “field”, and “textbook”. Based on these keywords, the topic was defined as “curriculum and implementation of invention and intellectual property education”. The presence of curriculum- and school-related terms suggests that this line of research addresses textbook content, curricular positioning, and school-level implementation of invention and intellectual property education.

The top keywords for Topic 5 included “school”, “invention education center”, “creativity”, “tools”, “digital”, “environment”, and “system”. Based on these keywords, the topic was defined as “educational environments, infrastructure, and tools for invention education.” It demonstrates research trends related to invention education infrastructure, including invention education centers and digital environments.

5. Conclusion and Implications

This study analyzed trends and structural characteristics of recent research in secondary school invention education through topic modeling based on 80 articles published in KCI-indexed domestic academic journals between 2020 and 2025.

The analysis identified five dominant topic clusters. Topic 1, labeled as AI- and maker-based problem-solving program development, reflects a concentration on the design of instructional programs incorporating emerging technologies. Topic 2, labeled as teaching-learning design, teachers, and policy support for invention education, included policy-related keywords but accounted for the smallest proportion. While this indicates active instructional innovation, the relative scarcity of keywords associated with long-term implementation, sustainability, or large-scale institutional application suggests that research attention remains largely development-oriented rather than longitudinally evaluative.

Topic 3, identified as entrepreneurship and startup competency education, demonstrates the integration of entrepreneurial perspectives within invention education. Topic 4, labeled as, curriculum and implementation of invention and intellectual property education, indicates that curricular and institutional discussions are present within the field. However, compared with program-level research, curriculum-centered studies appear relatively less dominant, suggesting a structural imbalance between instructional innovation and formal curricular integration. This suggests that future invention education should move beyond program development and short-term effectiveness evaluation, and expand toward organic integration with entrepreneurship education, systematic alignment with the formal curriculum, and stronger sustainable implementation and institutional support at the school level.

Topic 5, labeled as educational environments, infrastructure, and tools for invention education, highlights the need to continuously incorporate emerging technologies into invention education infrastructure, including invention education centers and digital environments, and suggests the importance of strengthening such infrastructure and effectively integrating it into teaching and learning.

Overall, the findings portray a research landscape centered on inventive thinking in which maker, entrepreneurship, and intellectual property elements coexist within secondary school invention education studies. However, the emphasis on program development and short-term effectiveness evaluation highlights the need to strengthen longitudinal research designs, teacher capacity development, and school-level institutional integration in future studies. A more balanced integration of these domains may contribute to strengthening the coherence and developmental maturity of scholarship in secondary invention education.

References

Canedo, M. (1989). The inventive thinking curriculum project. DIANE Publishing.

Dalela, S., & Ahmed, M. S. (2024). Systematic review of invention education research landscape: State of the discipline and future directions. Frontiers in Education, 9, 1284442.

Kim, B., & Park, C. (2025). The relationship between career-linked invention education experience satisfaction and career and learning management attitudes: Focusing on general high school students. Korea Society of International Relations and Education, 5(2), 247-263.

Kim, H. (2022). Development of a non-face-to-face remote invention education program for elementary school students. Journal of Convergence Science, Technology, and Society, 1(2), 21-25. Korean Intellectual Property Office. (2017). Act on the promotion and support of invention education.

http://www.law.go.kr

Korean Intellectual Property Office. (2022). Vision 2030 for intellectual property human resource development (Second master plan for invention education). https://www.kipo.go.kr

Lee, E. (2015). Research trends and issues of invention education in early childhood, elementary and secondary education : Focussing on Journals in Korea. Journal of Teacher Education, 31(3), 333-356.

Lee, K., & Tae, J. (2013). Development and application effects of creative problem-solving invention program of focusing on team project. The Journal of The Korean Society for the Gifted and Talented, 12(3), 95-116.

Michael, A. G., Salmon, K. R., Testorf, M. E., Morrone, M., Bass, K. M., & Faletra, P. (2021). STEM lab on a kitchen table: An investigation of remote student-driven problem-based research. Journal of STEM outreach, 4(2), 10-15695.

Moon, D. (2022). A study on the changes of primary and secondary invention education policy in Korea. Journal of the Korean Technology Education Association, 22(1), 161-178.

Scharon, C. J., Phillips, A., & Jones-Davis, D. (2024). The mind of a maker: a learning framework for a continuum of K-12 invention education. Frontiers in Education, 9, 1352274.

Hsu, T. C., & Ogata, H. (Eds.). (2026). *Proceedings of the 10th International Conference on Artificial Intelligence, Computational Thinking and STEM Education* .Kumamoto: The Asia-Pacific Society for Computers in Education.

An Epistemic Gap in Computational Thinking and AI Integration among Pre-Service Biology Teachers: A Qualitative Needs Analysis

Suhanna ZAINUDIN[1], Hidayah MOHD FADZIL[2*], Edy Hafizan MOHD SHAHALI[3], Mohd Izwan OTHMAN[4]
[1,2,3] Faculty of Education, Universiti Malaya, Malaysia
[4] City University, Malaysia
* hidayahfadzil@um.edu.my

Abstract: *The integration of Computational Thinking (CT) and Artificial Intelligence (AI) has become increasingly significant in education, particularly within science disciplines such as Biology. While both CT and AI are recognised as essential competencies, existing research has largely focused on technological exposure rather than how pre-service teachers conceptually understand and integrate these domains. This study examines this issue through the lens of an epistemic gap, defined as a misalignment between what individuals do and what they understand. A qualitative needs analysis was conducted involving five pre-service Biology teachers and two experts. Data were collected through semi-structured interviews and analysed using reflexive thematic analysis. The findings reveal that participants implicitly demonstrate CT practices such as decomposition and structured problem-solving, yet lack the ability to explicitly articulate these processes. In addition, CT is often perceived as dependent on technology, reflecting a limited understanding of its broader cognitive nature. Similarly, AI is primarily used as a functional tool for task completion, with limited awareness of underlying processes such as data classification and modelling. These findings indicate a gap between procedural engagement and conceptual understanding, which constrains the pedagogical integration of CT and AI. To address this issue, the study suggests that instructional approaches should explicitly connect CT concepts with AI processes. Engaging pre-service teachers in simple machine learning activities may support deeper conceptual understanding and facilitate more meaningful integration in teaching contexts. The study contributes by reframing CT-AI integration as an issue of epistemic alignment in teacher education.*

Keywords: Computational Thinking, Artificial Intelligence, pre-service teachers.

1. Introduction

The increasing integration of Computational Thinking (CT) and Artificial Intelligence (AI) in education has positioned both as essential competencies, particularly within science disciplines. CT is widely recognised as a problem-solving framework involving processes such as decomposition, abstraction, and algorithmic reasoning, extending beyond programming to support structured thinking across contexts (Wing, 2006, 2011; Voogt et al., 2015). In parallel, AI enables data-driven inquiry, pattern recognition, and predictive modelling, offering new ways of constructing knowledge in education (Holmes et al., 2022; Zawacki-Richter et al., 2019). Recent frameworks further emphasise that AI literacy in education requires not only technical use but also conceptual and ethical understanding of how AI systems function (UNESCO, 2023). In Biology, where classification, modelling, and analysis of complex systems are central, the integration of CT and AI has strong potential to enhance scientific understanding (Weintrop et al., 2016).

However, existing research has largely emphasised technological exposure and skill development, with less attention given to how pre-service teachers conceptually understand CT and AI (Angeli & Giannakos, 2020; Yadav et al., 2017). This distinction is critical, as the ability to use tools does not necessarily reflect conceptual understanding. From this perspective, CT can be understood as a way of thinking, while AI represents a data-driven system through which knowledge is generated. Importantly, emerging work suggests that the rise of machine learning highlights a distinction between rule-based and data-driven approaches to problem-solving, reshaping how CT is enacted in practice (Shamir & Levin, 2022). Similarly, recent studies highlight that while individuals are increasingly using AI applications, they often lack understanding of the underlying concepts and technologies (Ng et al., 2021). This gap may extend to pre-service teachers, where the use of AI tools does not necessarily translate into pedagogical or conceptual integration. Meaningful integration therefore requires an alignment between these ways of thinking and knowing.

When this alignment is not established, a gap emerges between what individuals do and what they understand. This study refers to this condition as an epistemic gap, where pre-service teachers engage in computational practices and use AI tools without the conceptual grounding needed to explain, transfer, or apply them in teaching contexts. Accordingly, this study examines how CT and AI are understood and enacted among pre-service Biology teachers, with the aim of explaining the conditions that shape readiness for meaningful CT-AI integration.

2. Methodology

This study employed a qualitative design to examine how pre-service Biology teachers understand and engage with Computational Thinking (CT) and Artificial Intelligence (AI). A needs analysis approach was adopted to explore participants' existing conceptions, misconceptions, and readiness prior to instructional intervention.

Participants were selected using purposive sampling to ensure alignment with the study focus (Creswell & Poth, 2018). The study involved five pre-service Biology teachers enrolled in a teacher education programme at a public university. In addition, two experts in Computational Thinking with experience in teacher professional development were included to provide contextual perspectives that supported the interpretation of the findings.

Data were collected through semi-structured interviews, which allowed participants to describe their experiences and perspectives while maintaining consistency across the dataset (Kvale & Brinkmann, 2009). All interviews were audio-recorded, transcribed verbatim, and analysed using reflexive thematic analysis (Braun & Clarke, 2006, 2019).

The analysis followed an inductive process. Initial open coding was conducted to identify meaningful data segments, followed by iterative grouping of codes into broader themes. Throughout the process, coding and theme development were continuously reviewed to ensure clarity and coherence. Given the focused scope of the study, emphasis was placed on depth of interpretation, with data considered sufficient when no new patterns emerged across participants.

3. FINDING

The analysis revealed a consistent pattern indicating a gap between participants' practices and their conceptual understanding of Computational Thinking (CT) and Artificial Intelligence (AI). This pattern is presented through three themes.

3.1. Implicit Computational Thinking without Explicit Understanding

All participants demonstrated elements of CT in their academic work, particularly in organising tasks, breaking down problems, and planning steps. However, these practices were not

recognised as part of CT. For example, one participant explained, "I usually break the assignment into several smaller parts first… before completing everything" (P1). Despite demonstrating such processes, several participants were unable to define CT or identify its components. This suggests that CT is present in practice but not understood conceptually.

3.2. Technology-Centric View of Computational Thinking

Participants commonly associated CT with the use of digital tools or computers. CT was described as something that requires technology, rather than a way of thinking that can be applied across contexts. As one participant noted, "If we don't use computers or technology, I'm not really sure how to relate it to CT" (P3). Expert participants also highlighted this misconception, explaining that CT is often reduced to coding or tool usage. This reflects a limited understanding of CT as a cognitive process.

3.3. Use of AI as a Tool without Pedagogical Understanding

All participants reported using AI tools for tasks such as generating ideas, summarising information, and searching for resources. One participant stated, "I usually use ChatGPT to get initial ideas or a draft before I refine it myself" (P3). However, this use remained task-oriented. Participants did not demonstrate understanding of how AI systems function and expressed low confidence in using AI for teaching. AI was often perceived as complex and technical, leading to a separation between usage and classroom application.

Overall, the findings show that while participants engage with CT and AI in practice, their understanding remains limited, resulting in a gap between what they do and what they are able to explain or apply in teaching contexts.

4. DISCUSSION AND CONCLUSION

This study reveals a gap between pre-service Biology teachers' practices and their conceptual understanding of Computational Thinking (CT) and Artificial Intelligence (AI). Although participants demonstrated elements of CT through structured problem-solving and reported frequent use of AI tools, these practices were not supported by explicit conceptual awareness. This suggests that exposure to technology alone does not necessarily lead to meaningful understanding or pedagogical readiness.

Consistent with prior research (Wing, 2006; Voogt et al., 2015), participants were able to perform computational processes such as decomposition and structured reasoning, yet were unable to articulate them. Similarly, their use of AI was largely instrumental, focusing on task completion rather than understanding underlying processes such as data classification and modelling. This aligns with findings that highlight limited pedagogical grounding and awareness of AI among educators (Zawacki-Richter et al., 2019; Holmes et al., 2022).

These findings point to an epistemic gap between what pre-service teachers do and what they understand, limiting the transfer of CT and AI into teaching contexts. To address this, instructional approaches should emphasise explicit connections between CT concepts and AI processes. Engaging pre-service teachers in simple machine learning tasks such as organising data, defining categories, and interpreting outputs can make visible how CT operates within data-driven systems, consistent with the view that machine learning extends computational thinking (Shamir & Levin, 2022).

This study contributes by reframing CT-AI integration as an issue of epistemic alignment, highlighting the need for conceptually grounded and discipline-specific instructional approaches in teacher education. Such approaches may inform the design of teacher training modules that explicitly connect computational thinking with AI processes in subject-specific contexts.

Acknowledgements

The authors gratefully acknowledge the Ministry of Education Malaysia and the University of Malaya for their continuous support and encouragement.

References

Angeli, C., & Giannakos, M. (2020). Computational thinking education: Issues and challenges.
Computers in Human Behavior, 105, 106185.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101.

Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis. Qualitative Research in Sport, Exercise and Health, 11(4), 589–597.

Creswell, J. W., & Poth, C. N. (2018). Qualitative inquiry and research design: Choosing among five approaches (4th ed.). SAGE Publications.

Holmes, W., Bialik, M., & Fadel, C. (2019). Artificial intelligence in education promises and implications for teaching and learning. Center for Curriculum Redesign.

Kvale, S., & Brinkmann, S. (2009). InterViews: Learning the craft of qualitative research interviewing (2nd ed.). SAGE Publications.

Ng, D. T. K., Leung, J. K. L., Chu, K. W. S., & Qiao, M. S. (2021). AI literacy: Definition, teaching, evaluation and ethical issues. Proceedings of the association for information science and technology, 58(1), 504-509.

Shamir, G., & Levin, I. (2022). Teaching machine learning in elementary school. International Journal of Child-Computer Interaction, 31, 100415.

UNESCO. (2023). *AI competency framework for teachers*. UNESCO.

Voogt, J., Fisser, P., Good, J., Mishra, P., & Yadav, A. (2015). Computational thinking in compulsory education: Towards an agenda for research and practice. Education and information technologies, 20(4), 715-728.

Weintrop, D., Beheshti, E., Horn, M., Orton, K., Jona, K., Trouille, L., & Wilensky, U. (2016). Defining computational thinking for mathematics and science classrooms. *Journal of science education and technology*, *25*(1), 127-147.

Wing, J. M. (2006).Computational thinking. Communications of the ACM, 49(3), 33–35.

Wing, J. (2011). Research notebook: Computational thinking—What and why. The link magazine, 6(1), 20-23.

Yadav, A., Stephenson, C., & Hong, H. (2017). Computational thinking for teacher education. Communications of the ACM, 60(4), 55-62.

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education. International Journal of Educational Technology in Higher Education, 16(1), 39.

Hsu, T. C., & Ogata, H. (Eds.). (2026). *Proceedings of the 10th International Conference on Artificial Intelligence, Computational Thinking and STEM Education* .Kumamoto: The Asia-Pacific Society for Computers in Education.

Developing an LLM-Based System to Support Teachers' Task Design for Students' Deep Understanding

Satomi SHIBA[1*], Hirohisa ODA[2] & Yuri UESAKA[3]
[1] Institute of Science Tokyo, Japan
[2] University of Shizuoka, Japan
[3] The University of Tokyo, Japan
* shiba.s.bacb@m.isct.ac.jp

Abstract: *Promoting students' deep understanding is a central goal in school education, and the learning tasks provided by teachers play a crucial role. However, designing learning tasks that extend beyond typical tasks and explicitly promote deep understanding remains challenging for teachers. This study proposes a large language model (LLM)-based system that supports teachers' task design through collaboration with generative AI. The system includes the following features: research-informed scaffolding to focus on students' deep understanding, task design based on points of difficulty in students' understanding, and task idea generation from multiple perspectives to broaden teachers' thinking. User evaluations were conducted with three teachers from Japanese school. Results indicated that the system was perceived as useful in reducing the cognitive and time-related demands of task design. Teachers particularly valued the generation of diverse task ideas and the interactive process of refining tasks with the system. At the same time, some areas for improvement were identified, such as usability and visual outputs. These findings suggest the potential of generative AI in supporting teachers' instructional practice, while highlighting the need for further system refinement and more rigorous evaluation with larger samples.*

Keywords: Teacher-AI collaboration, Task design, LLM-based system, Deep understanding, K-12 education

1. Introduction

In Japanese school education, ensuring deep understanding for all students is identified as a central goal of subject learning, as it serves as a foundation for competencies such as critical thinking and creative problem solving (MEXT, 2017). The concept of deep understanding has been examined in educational psychology, which involves understanding the reasons underlying procedures, the meanings of concepts, or the connections among knowledge (Bransford, Brown, & Cocking, 2000; Crooks & Alibali, 2014). In contrast, shallow understanding refers to fragmented knowledge that cannot be applied or transferred.

Promoting students' deep understanding requires consistent support in lessons, with particular emphasis on learning tasks provided by teachers. However, typical tasks found in textbooks tend to focus on procedural application or basic knowledge checks. This requires teachers to design tasks that explicitly promote and assess deep understanding. In practice, even when teachers can conceptualize students' deep understanding, they often face difficulty translating it into task design (Tchoshanov & Monarrez, 2024). Although research on teachers' task design remains limited, our study suggests that considering points of difficulty in students' understanding can support effective task design (Uesaka et al., 2024).

Generative AI systems based on large language models (LLMs) have attracted increasing attention as tools to support teachers' instructional work (e.g., Celik et al., 2022; Hu et al., 2025). However, research on how these systems can enrich teachers' practice in substantive ways remains limited. In the context of task design, a key issue is how teachers can effectively engage with AI as a collaborative partner when designing tasks in their daily practice, moving beyond simply accepting AI-generated outputs.

This study aims to develop an LLM-based system to support teachers' task design for promoting students' deep understanding. This paper presents the system and reports findings from a preliminary user evaluation conducted with teachers in Japanese schools.

2. System Development

2.1. Core Features of the System

The proposed system is a web-based application that utilizes LLMs to support teachers' task design in school education. Rather than automatically generating ready-made tasks, the system is aimed at supporting collaborative task design between teachers and generative AI. To achieve this goal, the system incorporates the following three core features.

Research-informed scaffolding for task design. The system incorporates research-based knowledge on learning and understanding to orient teachers' thinking toward students' deep understanding during task design. By presenting simulated student responses at different levels of understanding, the system encourages teachers to reflect on how students understand the learning topic.

Task design based on points of difficulty in understanding. Considering points of difficulty in students' understanding, referred to in this study as tsumazuki-points, is regarded as an effective task design strategy (Uesaka et al., 2024). In the system, these points are suggested by an LLM or provided by teachers and are used as perspectives for task design.

Generating task ideas from multiple perspectives. To broaden teachers' thinking in task design, the system presents multiple task ideas rather than proposing a single task.

These ideas are grounded in perspectives derived from case studies of experienced teachers (Uesaka et al., 2024), while allowing the LLM flexibility to generate additional ideas beyond typical tasks that focus mainly on procedural application or basic knowledge checks. This approach encourages teachers to engage in comparing, selecting, and refining alternative ideas.

2.2. System Workflow

The proposed system follows a workflow consisting of eight steps (Figure 1). The system is accessible via a web browser on PCs, tablets, and smartphones. The backend is implemented in Python, and the user interface is built with Gradio. Multi-agent coordination and self-reflection are implemented using LangGraph, and the system employs multiple open-weight LLMs (e.g., DeepSeek-R1, Kimi K2 Thinking).

Step 1: Specifying a topic. The user specifies a learning topic or a typical task. The system estimates the subject area and school grade, and generates a typical task and its solution when only a learning topic is provided.

Step 2: Simulating students' understanding. The system simulates a dialogue involving two students and one teacher. One student represents shallow understanding based on rote memorization and repetitive practice, while the other has lower academic performance but raises questions that lead to deeper understanding.

Step 3: Identifying a *tsumazuki-point*. The user inputs a *tsumazuki-point* or selects one from four candidates generated by the system from different perspectives.

Step 4: Suggesting task ideas. The system presents seven distinctive task ideas, from which the user selects one. These ideas are generated based on the specified *tsumazuki-point*, considering response formats (e.g., multiple-choice) and task design perspectives (Uesaka et al., 2024). To support this process, the system adopts a multi-agent architecture consisting of two LLM roles: an idea generator and a critic. The idea generator produces multiple task ideas with explanations of their instructional goals, and the critic selects and refines appropriate ones.

Steps 5–6: Generating an initial task. Based on the specified *tsumazuki-point* and the selected task idea, the system generates and presents an initial task.

Steps 7–8: Updating the task. Task refinement is supported through two simulations: expert review by another LLM and a problem-solving simulation by the two students introduced in Step 2, both of which generate suggestions for revision. The user can also input additional revision comments.

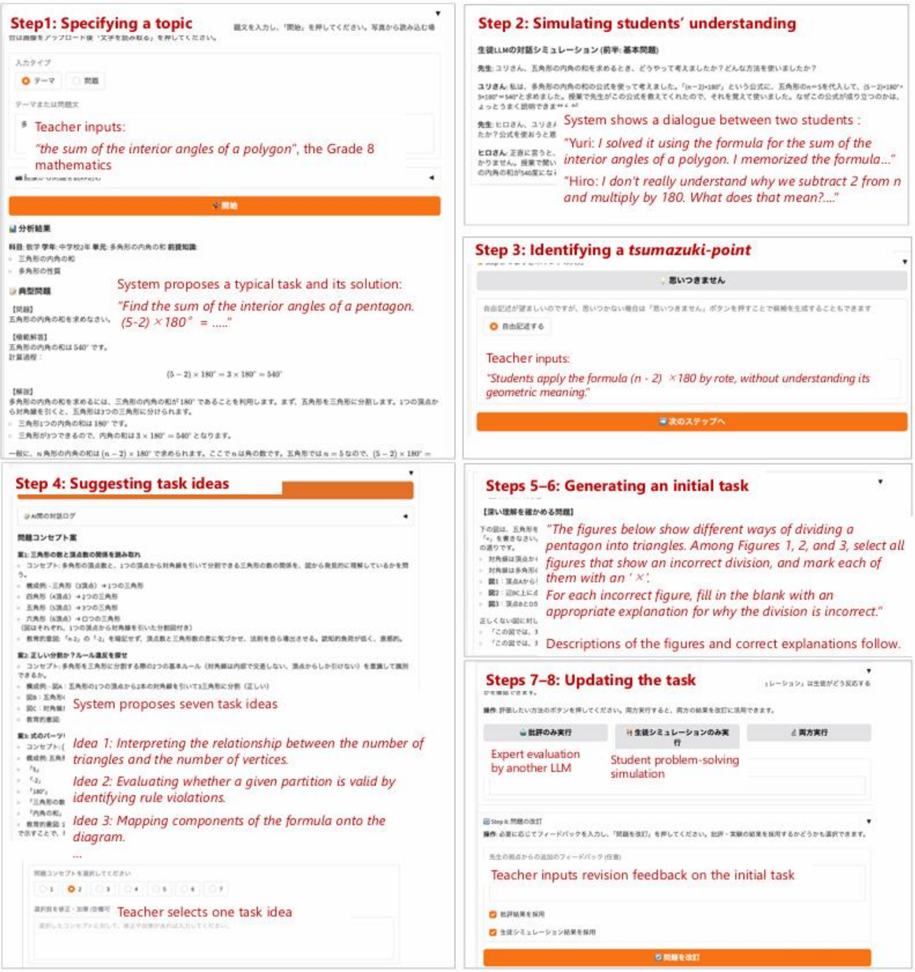


Figure 1. Flow of the proposed system.

3. User Evaluation

Three teachers with experience in task design participated in a one-hour online session: one elementary school teacher (female, T1), one high school math teacher (male, T2), and one high school science teacher (male, T3). Following system use, a semi-structured interview was conducted, focusing on overall impressions, perceived usefulness, and areas for improvement. In addition, participants completed a short questionnaire with six items rated on a five-point Likert scale: usefulness, interest, difficulty of use, intention to use, intention to recommend, and need for improvement.

The results indicated overall positive evaluations. All three teachers rated the system as useful and interesting, with strong intentions to use and recommend it (M = 5.0 for all items), while perceived difficulty of use was low (M = 2.0). In the interviews, all teachers reported that the system reduced the cognitive and time demands of task design. In particular, the suggestion of task ideas (Step 4) was highly valued. For example, T2 noted that it provided ideas that might not have occurred to him and helped clarify vague thoughts. T3 reported that he enjoyed designing tasks while commenting on the system's outputs.

Several areas for improvement were identified, as also indicated by the questionnaire results (need for improvement, M = 4.0). T1 and T2 pointed out that system outputs were often lengthy, making them time-consuming to read and obscuring key points. T3 reported difficulty in formulating inputs and expressing intentions clearly due to limited familiarity with generative AI systems. For the elementary mathematics topic, some generated tasks did not adequately reflect the curricular progression of learning content. In addition, all teachers suggested that the system should present diagrams or visual representations.

3. Discussion

This study presented an LLM-based system that supports teachers' task design for promoting students' deep understanding through interaction with generative AI under research-informed scaffolding. The user evaluation indicated that the system reduced cognitive and time demands and helped clarify and expand teachers' thinking in task design. Further system development will focus on usability, including visualization features and more concise system-generated outputs. Efforts will also focus on enhancing the quality of generated task ideas by accounting for curricular progression and refining simulations of students' understanding, thereby supporting teachers in better identifying tsumazuki-points.

However, these findings should be interpreted as preliminary, as the evaluation was conducted with a small sample and relied primarily on subjective measures. Future research should involve larger samples and more rigorous user evaluations, including analyses of generated task quality and users' cognitive processes during interaction with the system.

Another important direction is to investigate how the system can contribute to teachers' task design skills and its effective use in classroom practice.

Acknowledgements

This study was supported by Grants-in-Aid for Scientific Research (JSPS KAKENHI, Number 23H00065). We would like to thank the teachers who supported the system development.

References

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). How people learn (Vol. 11). Washington, DC: National academy press.

Celik, I., Dindar, M., Muukkonen, H., & Järvelä, S. (2022). The promises and challenges of artificial intelligence for teachers: A systematic review of research. TechTrends, 66(4), 616-630.

Crooks, N. M., & Alibali, M. W. (2014). Defining and measuring conceptual knowledge in mathematics. Developmental review, 34(4), 344-377.

Hu, B., Zhu, J., Pei, Y., & Gu, X. (2025). Exploring the potential of LLM to enhance teaching plans through teaching simulation. npj Science of Learning, 10(1), 7.

Ministry of Education, Culture, Sports, Science and Technology. (2017). Course of study for elementary school. https://www.mext.go.jp/content/20230120-mxt_kyoiku02-100002604_01.pdf

Tchoshanov, M., & Monarrez, A. (2024). Analyzing teacher-designed tasks through the lens of procedural and conceptual knowledge. Canadian Journal of Science, Mathematics and Technology Education, 24(1), 23-38.

Uesaka, Y., Sakaguchi, T., Mizuno, Y., Furuya, K., Terui, Y., Kurosawa, T., Miyawaki, Y., Shiba, S., Uchida, N., Liu, M., & Zhang, E. (2024). Exploring design principles for learning tasks that promote deep understanding [Poster presentation]. Japan Society for Educational Technology Autumn National Conference 2024, Japan. (In Japanese)

Hsu, T. C., & Ogata, H. (Eds.). (2026). *Proceedings of the 10th International Conference on Artificial Intelligence, Computational Thinking and STEM Education* .Kumamoto: The Asia-Pacific Society for Computers in Education.

A Generative AI–Based Job Creation Career Education Program Grounded in Career Adaptability

Young Ji KIM[1], Jeongmin LEE[2*]
[1,2] Department of Educational Technology, Ewha Womans University, South Korea
*jeongmin@ewha.ac.kr

Abstract: *This study developed and conducted an initial evaluation of a generative AI-based job creation career education program grounded in career adaptability theory for university students. Using a developmental research design, the program was implemented with university students and evaluated through expert review and learner usability assessment. Findings suggest that AI-based job creation activities can support learners' perceived career adaptability by providing structured opportunities for future exploration, goal setting, role ideation, and feasibility testing. Expert and learner evaluations also identified the need to improve interaction continuity, prompt coherence, and pre-instructional guidance.*

Keywords: Career Adaptability, Job Creation, Generative AI, University students, Program Design

1. Introduction

Rapid technological shifts and AI integration have rendered traditional linear career planning inadequate, leaving university students to navigate heightened career uncertainty (World Economic Forum, 2025). To address this, contemporary career education must prioritize "career adaptability", a psychosocial competence comprising concern, control, curiosity, and confidence (Savickas & Porfeli, 2012). Despite its developmental importance, current university interventions often remain limited to static job-search strategies, failing to provide learners with personalized, agentic career design experiences (Kim & Jang, 2019).

To address these limitations, this study proposes "job creation", the iterative process of constructing new professional roles by synthesizing individual strengths with contextual opportunities, as a core learning activity (Korean Employment Information Service [KEIS], 2019). However, standard educational practices often fall short in facilitating the personalized contexts required to design new occupational roles (Green et al., 2020), indicating a need for more robust, higher-order learning experiences.

Generative AI offers transformative potential for this purpose by enabling personalized feedback, real-time scenario generation, and feasibility testing (Mehlan et al., 2025). While AI is frequently applied to basic content tasks, its structured use in designing career programs that guide students through pathway exploration and role construction remains limited (Sekli et al., 2024). Consequently, this study designs and evaluates a generative AI-integrated program aimed at fostering university students' career adaptability, validated through expert content assessment and learner-centered usability testing. The research questions guiding this study are as follows:

1. What are the core design principles of a generative AI–based job creation career education program grounded in career adaptability theory?
2. How do experts and learners evaluate the program's validity, usability, and perceived support for the four dimensions of career adaptability, namely concern, control, curiosity, and confidence?

2. Literature Review

Within Career Construction Theory, career adaptability acts as a psychosocial resource that enables individuals to manage career transitions in volatile work environments (Savickas, 2019). Rather than mere conformity, it is an active process of designing meaningful lives and constructing new vocational forms (Merino-Tejedor et al., 2025). Savickas and Porfeli (2012) defined this construct through four interconnected dimensions: concern (future preparedness), control (self-determination), curiosity (exploration), and confidence (resilience). Beyond employment readiness, adaptability fosters life satisfaction and proactive engagement in university students (Chen et al., 2020; Gao et al., 2019).

Crucially, adaptability is a malleable competence that can be effectively enhanced through structured educational interventions (Kim, 2022; Wetstone & Rice, 2023).

Job creation is the active process of generating new occupational roles by integrating individual competencies with contextual opportunities, rather than merely adapting to existing job markets (KEIS, 2019). As technological shifts reshape the labor landscape, university students must move beyond traditional major-based preparation toward proactive, agentic career design (Frey & Osborne, 2017). This process functions as a practical enactment of career adaptability: "concern" facilitates envisioning emerging possibilities, "control" aids goal-setting, "curiosity" drives divergent exploration of role combinations, and "confidence" supports iterative prototyping (Savickas & Porfeli, 2012). Despite its potential for enhancing self-efficacy, many existing programs remain limited by static, short-term formats that lack personalized feedback and clear connections between ideation and execution (Kim et al., 2017). There is a critical need for more structured, AI-supported educational approaches to systematically guide learners through this design process (Zhang et al., 2025).

Generative AI acts as an intelligent, interactive tutor that provides personalized scaffolding, real-time simulation, and iterative feedback to support higher-order thinking (Ilgun Dibek, 2024; Wang et al., 2024). These capabilities are particularly effective for "job creation" education, which requires navigating open-ended tasks and high uncertainty.

Through conversational dialogue, generative AI facilitates divergent exploration and feasibility testing, allowing students to tailor career scenarios to their unique values and skills (Chen et al., 2022). Furthermore, as a co-designer, AI fosters learner agency by prompting reflection and role construction rather than providing static answers (Duan & Wu, 2024). To mitigate challenges regarding information reliability and prompt literacy, effective integration requires structured instructional guidance that emphasizes AI as a tool for creative agency rather than an automated solution provider (Bayaga, 2025). This study addresses the need for empirical research on generative AI as a learner-centered tool to support career adaptability.

3. Method

This study adopted a developmental research design to develop and conduct an initial evaluation of a generative AI–based job creation career education program grounded in career adaptability theory for university students. The research followed a multi-phase procedure: experts in educational technology, computer education conducted content validation during the development phase, while 10 university students from Korean institutions participated in the evaluation phase. Participants were recruited via convenience sampling from online university communities and provided electronic informed consent regarding the study’s purpose, data usage, anonymity, and voluntary withdrawal rights.

The program was developed using the ADDIE model, with the analysis phase drawing upon literature from RISS, KCI, and Google Scholar to integrate Savickas’s four dimensions of career adaptability(concern, control, curiosity, and confidence) into career creation activities. Structured as a two-hour, single-session intervention, the program used a constructive alignment approach to link these dimensions to specific objectives, activities, and AI-driven agent tools. Learners were scaffolded through a process where concern was fostered via future exploration simulations, control was strengthened through structured goal setting, curiosity was promoted through divergent AI-led ideation, and confidence was supported through iterative feasibility analysis and prototype development.

This study employed validated instruments to assess key variables, with all items rated on a 5-point Likert scale. Career adaptability was measured using the Korean version of the Career Adapt-Abilities Scale–Short Form (K-CAAS-SF), validated by Tak et al. (2015) and further refined by Kim and Ko (2020). Career decision-making self-efficacy was assessed based on the framework established by Yoon (2020). Finally, program satisfaction and usability were evaluated using the scale developed by Shim and Park (2024).

During implementation and data analysis, expert evaluation and learner evaluation were conducted separately. Experts assessed the program's theoretical alignment, content appropriateness, AI feedback quality, and real-world feasibility using a Likert-type scale, alongside qualitative suggestions for technical refinement. Subsequently, university students participated in a usability evaluation, where quantitative data on career adaptability, career decision-making self-efficacy, and program satisfaction were collected via Likert scales. Open-ended written feedback was also analyzed to identify recurring strengths, limitations, and revision needs.

4. Findings & Conclusion

The expert review confirmed the academic value of the program and the novelty of the AI agent design, while highlighting the need to refine the scope of activities and the use of diagnostic tools. Accordingly, the program was revised to focus on supporting the early stages of job creation, prioritizing motivation and self-exploration, and integrating the Korea Collegiate Essential Skills Assessment (K-CESA) to provide tailored competency development recommendations.

Based on the development and evaluation process, the program was organized around four design principles: theory-aligned scaffolding based on the four dimensions of career adaptability, learner-agency-centered use of generative AI, diagnostic personalization through competency-based feedback, and iterative feasibility testing for emerging occupational roles.

The usability evaluation indicated generally favorable learner perceptions of the program's career-related value, with mean scores of 4.08 for career adaptability and 3.87 for career decision-making self-efficacy. However, usability and satisfaction showed a relatively lower mean score of 3.66, likely reflecting participants' focus on technical interactions with the AI agent rather than the program's overall instructional design. Qualitative feedback highlighted the program's novelty and value in supporting creative job creation, while also identifying limitations such as interaction interruptions, contextually inappropriate responses, lack of coherence, and repetitive information. These findings suggest the need to improve pre-instructional guidance, refine question structures, and enhance AI prompts.

Despite these contributions, several limitations remain. The small sample size and

short-term, single-session design limit the generalizability of the findings and preclude robust causal inference regarding the program's effects. In addition, technical instability in generative AI reduced interaction continuity and learner immersion, while structured AI feedback may paradoxically constrain exploratory thinking. Repetitive information also led to cognitive fatigue, indicating the need to optimize interaction coherence and diversity. Future research should refine the agent's questioning logic and system architecture, incorporate more sophisticated prompts that support deeper reflection on emerging career roles, and conduct longitudinal studies extending to practical outputs, such as career proposals and their real-world implementation, to further examine the program's educational value and sustainability.

Overall, this study provides preliminary design evidence that a generative AI–based job creation career education program can support learners' exploration, agency, and confidence when structured around the four dimensions of career adaptability. However, the results should be interpreted cautiously due to the small convenience sample, the short-term single-session format, and the technical instability observed during AI interaction.

Acknowledgements

This study was supported by Grants-in-Aid for Scientific Research (JSPS KAKENHI, Number 23H00065). We would like to thank the teachers who supported the system development.

References

Bayaga, A. (2025). Leveraging AI-enhanced and emerging technologies for pedagogical innovations in higher education. Education and Information Technologies, 30(1), 1045–1072.

Chen, H., Fang, T., Liu, F., Pang, L., Wen, Y., Chen, S., & Gu, X. (2020). Career Adaptability Research: A Literature Review with Scientific Knowledge Mapping in Web of Science. International Journal of Environmental Research and Public Health, 17(16), 5986.

Chen, J., Chen, Y., Ou, R., Wang, J., & Chen, Q. (2022). How to use artificial intelligence to improve entrepreneurial attitude in business simulation games: implications from a quasi-experiment.

Frontiers in psychology, 13, 856085.

Duan, J., & Wu, S. (2024). Beyond traditional pathways: Leveraging generative AI for dynamic career planning in vocational education. International Journal of New Developments in Education, 6(2), 24-31.

Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? Technological Forecasting and Social Change, 114, 254–280.

Gao, X., Xin, X., Zhou, W., & Jepsen, D. M. (2019). Combine your “will” and “able”: career adaptability’s influence on performance. Frontiers in Psychology, 9, 2695.

Green, Z. A., Noor, U., & Hashemi, M. N. (2020). Furthering Proactivity and Career Adaptability Among University Students: Test of Intervention. Journal of Career Assessment, 28(3), 402-424.

Ilgun Dibek, M., Sahin Kursad, M., & Erdogan, T. (2024). Influence of artificial intelligence tools on higher order thinking skills: a meta-analysis. Interactive Learning Environments, 33(3), 2216–2238.

Kim, J. (2022). Development and effectiveness verification of an online career adaptability program for undergraduate students. Frontiers in Psychology, 13, 857111.

Kim, J., Choi, Y., & Choi, H. (2017). A basic study on the development of a job creation education program (Year 2): Development of a job creation education operation manual and contents. Korean Employment Information Service.

Kim, J., & Jang. E. (2019). Development and effectiveness verification of the tailored career group program for improving the career adaptability for university students. Korean Journal of Youth Studies, 26(9), 125-159. https://doi.org/10.21509/KJYS.2019.09.26.9.125

Kim, M., & Ko, E. (2020). Validation of the Korean version of the Career Adapt-Abilities Scale–Short Form (K-CAAS-SF). Korean Journal of Teacher Education, 36(4), 261–281. https://doi.org/10.14333/KJTE.2020.36.4.261

Korean Employment Information Service. (2019). Youth job creation program: Educational manual.

Mehlan, R., Hess, C., Stierstorfer, Q., & Schaaff, K. (2025). Personalized knowledge transfer through generative AI: Contextualizing learning to individual career goals. Arxiv, Advance online publication.

Merino-Tejedor, E., Serrano-Fernández, M. J., Boada-Cuerva, M., & Sora, B. (2025). Career adaptability and career construction as mediating variables between hardiness and vocational identity. International Journal for Research in Vocational Education and Training (IJRVET), 12(1), 127-150.

Savickas, M. L., & Porfeli, E. J. (2012). Career Adapt-Abilities Scale: Construction, reliability, and measurement equivalence across 13 countries. Journal of vocational behavior, 80(3), 661-673. https://doi.org/10.1016/j.jvb.2012.01.011

Savickas, M. L. (2019). Career construction theory. In J. A. Athanasou & H. N. Perera (Eds.), International handbook of career guidance (2nd ed., pp. 3-43). Springer.

Sekli, G. M., Godo, A., & Véliz, J. C. (2024). Generative AI solutions for faculty and students: A review of literature and roadmap for future research. Journal of Information Technology Education: Research, 23, 014.

Shim, H., & Park, I. (2024). Development and validation of a measurement tool for extracurricular program satisfaction in universities. Journal of University Teaching and Learning, 17(3), 111–

131. https://doi.org/10.23122/kactl.2024.17.3.005

Tak, J., Lee, E., & Lim, G. (2015). Validation of the career adaptability scale. Korean Journal of Industrial and Organizational Psychology, 28(4), 591–608.

Wang, S., Wang, F., Zhu, Z., Wang, J., Tran, T., & Du, Z. (2024). Artificial intelligence in education: A systematic literature review. Expert Systems with Applications, 252, 124167.

World Economic Forum. (2025). The future of jobs report 2025.

Wetstone, H., & Rice, K. G. (2023). Enhancing career adaptability among university students: An intervention study. Journal of Career Development, 50(6), 1279–1292.

Yoon, J. H. (2020). The effects of career decision-making self-efficacy on career preparation behavior and career adaptability among university students (Doctoral dissertation). Graduate School, University of Brain Education, Cheonan, Korea.

Zhang, J., Talib, M. A., & Wang, J. (2025). Enhancing career adaptability and career decision-making self-efficacy through career planning education: a quasi-experimental study. Frontiers in Psychology, 16, 1612768.

Evaluating the Impact of a Machine Learning Course on Pre-service Teachers in Hong Kong

Siu Cheung KONG[1*], Peilin LI[2]

[1] Department of Mathematics and Information Technology, The Education University of Hong Kong, Hong Kong SAR

[1,2] Artificial Intelligence and Digital Competency Education Centre, The Education University of Hong Kong, Hong Kong SAR

*sckong@eduhk.hk

Abstract: *This study examines the impact of a 6-hour robotics-based machine learning (ML) course on pre-service teachers' conceptual understanding and Technological Pedagogical Content Knowledge (TPACK) development. Thirty-nine pre-service teachers from diverse disciplinary backgrounds participated in the course. Quantitative analysis revealed significant improvements in participants' understanding of ML and TPACK development on teaching ML. Qualitative analysis of reflective writings and follow-up interviews highlighting three key themes: mastery of the three main machine learning paradigms, enhanced technical literacy regarding algorithms and neural networks, and a deeper understanding of the teaching of ML. Participants demonstrated increased pedagogical confidence and readiness to teach ML.*

Keywords: artificial intelligence, machine learning, pre-service teachers, professional development, TPACK

1. Introduction

Integrating machine learning (ML) effectively into curriculum teaching and learning can enhance students' subject-specific competencies (Priya et al., 2024). Effective use of ML requires a solid conceptual understanding of the technology (Kong & Hu, 2026). Therefore, offering professional development for teachers to deepen their understanding of ML is essential.

2. Literature Review

Existing research has examined various aspects of teaching machine learning (ML) in K–12 education (Martins et al., 2024). While many studies have investigated teachers' perceptions and professional development (Sanusi et al., 2023), most of this research has been limited to in-service teachers. There remains a notable lack of research into how pre-service teachers develop Technological Pedagogical Content Knowledge (TPACK) through hands-on robotics experiences (Bilbao-Eraña & Arroyo-Sagasta, 2025). To address this gap, this study seeks to evaluate the impact of a robotics-based ML course on pre-service teachers' conceptual understanding of ML and their development of TPACK for teaching capability enhancement. The research questions are as follows: (1) To what extent does a robotics-based ML course improve pre-service teachers' conceptual understanding of machine learning and their TPACK capability development in Hong Kong? (2) How do pre-service teachers perceive the impact of the machine learning course on their professional readiness?

3. Methodology

3.1. Machine Learning Course Design

The course introduces foundational ML concepts, algorithms, and neural networks through practical engagement with the AlphAI robot and software (Learning Robots, n.d.). The 6-hour, face-to-face course is structured around the

Attention-Engagement-Error-feedback-Reflection (AEER) pedagogical framework to support effective ML teaching and learning (Kong & Yang, 2024). The teaching covers K-Nearest Neighbors (KNN) algorithm (0.5 hour), artificial neural network (2.5 hours), supervised learning (2 hours), reinforcement learning and unsupervised learning (1 hours).

3.2. Participants

Participant recruitment was conducted via a university-wide email targeting Bachelor of Education students, in line with the study's focus on pre-service teacher development. A total of 39 students (N = 39) voluntarily registered and completed the 6-hour course. Participants represented all years of the programme: Year 1 (n = 9, 23.1%), Year 2 (n = 5, 12.8%), Year 3 (n = 5, 12.8%), Year 4 (n = 8, 20.5%), and Year 5 (n = 12, 30.8%). Most of the participants are majoring in the teaching of primary mathematics (35.9%), Chinese language (20.5%), and primary science (20.5%) and other disciplines (23.1%).

3.3. Research Methods

This research adopts a combination of quantitative and qualitative methods to evaluate the effectiveness of the ML course. The research instruments include (1) machine learning concept test, (2) TPACK survey, (3) reflective writing, and (4) follow-up interview. The ML concept test and TPACK survey were administered pre- and post-course. Reflective writing and follow-up interviews were implemented after completion of the course.

The machine learning concept test consists of 13 items designed to evaluate participants' conceptual understanding of ML. A sample question is provided: "Which of the following statements about machine learning is correct? A. Machines learn through logical reasoning. B. Machines cannot learn through rewards and punishment. C. Machines do not require data to learn. D. Machines can learn through labelled data."

The 15-item TPACK survey instrument is divided into four dimensions: Content Knowledge (CK), which includes 5 items (e.g., "I know how to train machine learning model using data"); Technological Content Knowledge (TCK), consisting of 3 items (e.g., "I know how to manipulate parameters on the interface of the software platform of the AlphAI robots to teach students input, hidden and output layers in the artificial neural network of deep learning"); Pedagogical Content Knowledge (PCK), with 4 items (e.g., "I know how to use Error-feedback effectively in Attention-Engagement-Error-feedback-Reflection pedagogy to teach students to learn key concepts of supervised learning in machine learning"); and Technological Pedagogical Content Knowledge (TPACK), comprising 3 items (e.g., "I know how to use the AlphAI robots and its software platform and the Attention-Engagement-Error-feedback-Reflection pedagogy effectively to teach machine learning and deep learning"). Participants rated their level of agreement from strongly disagree (1) to strongly agree (5) using a 5-point Likert scale.

Qualitative data was collected through reflective writing on their understanding of ML and follow-up interviews. The follow-up interview explored students' personal engagement and professional aspirations. Sample questions include "Would you like to teach what you have learned to your future students?"

3.4. Data Analysis

Prior to the main quantitative analysis, reliability testing yielded Cronbach's alpha of .71 for the ML concept test and .97 for the TPACK survey, representing acceptable and excellent internal consistency, respectively. Qualitative data, comprising reflective writing from 39 participants and five follow-up interviews, were analysed through thematic analysis.

4. Results

4.1. Quantitative data results

In response to RQ 1, the ML course significantly enhanced the pre-service teachers' conceptual understanding of ML and their TPACK development. For the ML concept test, a Shapiro-Wilk test was conducted to determine the normality of the data. The result indicated that the scores followed a normal distribution ($p = .273$). A paired-samples t-test showed a significant difference between pre-test ($M = 7.21$, $SD = 2.40$) and post-test ($M = 10.36$, $SD = 2.23$), $t(38) = 6.09$, $p < .001$. The Cohen's d was 0.976, indicating a large effect size and demonstrating the substantial effectiveness of the ML course.

Regarding the TPACK survey, normality for overall TPACK scores was assessed via a Shapiro-Wilk test, which indicated a non-normal distribution ($p < .05$). Given that the overall scores violated the assumption of normality, the non-parametric Wilcoxon signed-rank test was applied to all TPACK dimensions to maintain analytical consistency and ensure robust interpretation of the results. The results of the pre- and post-course comparisons across TPACK dimensions are presented in Table 1. The participants demonstrated significant improvements in every knowledge domain from pre-test to post-test ($p < .001$). These results indicate that the course was highly effective in enhancing pre-service teachers' TPACK development.

Table 1. *Comparison of Pre-test and Post-test Scores Across TPACK Dimensions (N=39)*

Dimensions	**Pre-test**	**Post-test**	***Z* value**	**Effect Size (*r*)**
	M (SD)	***M (SD)***		
CK	2.20 (1.06)	4.06 (0.45)	-5.20^{***}	.83
TCK	2.03 (0.18)	4.06 (0.08)	-5.24^{***}	.84
PCK	2.25 (0.19)	3.99 (0.09)	-4.98^{***}	.80
TPACK	2.23 (1.16)	3.99 (0.59)	-4.95^{***}	.79
Overall	2.18 (0.17)	4.03 (0.08)	-5.24^{***}	.84

$^{***}\ p < .001$

4.2. Qualitative data results

Regarding RQ2, the pre-service teachers' reflecting writing and follow-up interviews demonstrated significant conceptual development across three key themes: (1) mastery of the three main machine learning paradigms; (2) enhanced technical literacy regarding algorithms and neural networks; and (3) a deeper understanding of the teaching of ML. Furthermore, participants reported a high level of engagement, describing the course as an interesting experience, and expressed their confidence in teaching such knowledge.

Almost half of the participants reported that they had a clearer and critical understanding of the supervised, unsupervised, and reinforcement learning in ML. For example, participant 38 stated that "Supervised learning refers to machines learning from labelled data; reinforcement learning involves using rewards and punishments to guide the learning process; and unsupervised learning refers to machines learning from data that has no labels".

Besides, technical literacy of ML involves pre-service teachers' understanding and application of some algorithms and neural networks. Specifically, participant 35 reported that "Through this course, I gained practical experience in training robots to navigate clockwise using various techniques, including supervised learning, K-Nearest Neighbours (KNN), Artificial Neural Networks (ANN)—including hidden layers—overfitting, backpropagation, and reinforcement learning".

Participants' reflections also supported the practical application of the AEER pedagogical framework during the practical exercises. For instance, participant 12 emphasized "effective learning requires clear goals and provide the process with direction" (Attention), participant 1 indicated that "I believe learning occurs through hands-on experience rather than relying solely on textbook instruction" (Engagement), participant 5 thought that "allowing the robots to make mistakes and navigate its own solutions is an effective strategy to learn" (Error-feedback), and participant 14 perceived effective learning as "replacing passive repetition with active cognition and spaced practice, guided by specific objectives and continuous reflection and adjustment" (Reflection).

5. Implication and Conclusion

There are implications of the study. First, the study demonstrates that pre-service teachers, regardless of whether they have a science or non-science background, can develop an understanding of ML. Second, it shows that pre-service teachers have the capacity to teach ML when provided with training. Third, this study illustrates effective approaches for conducting teacher development in ML. In the future, we aim to investigate whether teachers with an understanding of ML can utilise it more effectively in their teaching practices.

Acknowledgements

The authors would like to acknowledge the funding from The Education University of Hong Kong with Project No. EdUHK 04A61: Using the Attention-Engagement-Error-Feedback-Reflection Pedagogy to Promote Learning-to-Learn Skills among Hong Kong Senior Primary Students / Junior Secondary Students through Understanding Machine Learning.

References

Bilbao-Eraña, A., & Arroyo-Sagasta, A. (2025). Fostering AI literacy in pre-service teachers: impact of a training intervention on awareness, attitude and trust in AI. Frontiers in Education (Lausanne), 10, Article 1668078. https://doi.org/10.3389/feduc.2025.1668078

Ertel, W. (2025). Introduction to artificial intelligence (3rd ed.). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-43102-0

Kong, S. C., & Yang, Y. (2024). Using the robot-assisted attention-engagement-error-feedback-reflection (AEER) pedagogical design to develop machine learning concepts and facilitate reflection on learning-to-learn skills: Evaluation of an empirical study in Hong Kong primary schools. In Proceedings of the 16th International Conference on Computer Supported Education - Volume 2 (pp. 155–162). Angers, France.

Kong, S. C., & Hu, W. (2026). Unleashing human potential: An artificial intelligence competency framework for K–12 education. Computers and Education: Artificial Intelligence, 10, 100556. https://doi.org/10.1016/j.caeai.2026.100556

Learning Robots. (n.d.). Teaching artificial intelligence in a simple way. https://www.learningrobots.ai/en

Martins, R. M., von Wangenheim, C. G., Rauber, M. F., & Hauck, J. C. (2024). Machine learning for all!—introducing machine learning in middle and high school. International Journal of Artificial Intelligence in Education, 34(2), 185–223. https://doi.org/10.1007/s40593-022-00325-y

Priya, S., Bhadra, S., Chimalakonda, S., & Venigalla, A. S. M. (2024). ML-Quest: A game for introducing machine learning concepts to K–12 students. Interactive Learning Environments, 32(1), 229–244. https://doi.org/10.1080/10494820.2022.2084115

Sanusi, I.T., Oyelere, S.S., Vartiainen, H. et al. (2023). A systematic review of teaching and learning machine learning in K–12 education. Education and Information Technologies, 28(5), 5967–5997. (2023). https://doi.org/10.1007/s10639-022-11416-7

Hsu, T. C., & Ogata, H. (Eds.). (2026). *Proceedings of the 10th International Conference on Artificial Intelligence, Computational Thinking and STEM Education* .Kumamoto: The Asia-Pacific Society for Computers in Education.

Uncovering Non-Linear Pathways of K-12 Teachers' AI Adoption Using Decision Trees

Tai-Ping HSU [1] & Ting-Chia HSU [2*]
[1,2] Department of Technology Application and Human Resource Development, National Taiwan Normal University
*ckhsu@ntnu.edu.tw

Abstract: *Although Artificial Intelligence (AI) literacy is critical for modern educators, a gap remains between theoretical knowledge and actual usage frequency, with traditional linear models often missing the complex, non-linear drivers of teacher behavior. To address this, a Classification and Regression Tree (CART) analysis was conducted with 155 K-12 teachers in Taiwan to identify hierarchical predictors of usage frequency using 40 specific literacy indicators. The results reveal a hierarchical structure where Critical Reflection acts as the primary gatekeeper for adoption. Pedagogical Efficiency emerged as the strongest overall predictor driving high-frequency use, while Self-Efficacy served as a key compensatory mechanism for teachers with developing reflective skills. Consequently, AI adoption is a non-linear process initiated by critical thought and sustained by efficiency. These findings support segmented professional development strategies that prioritize fostering critical reflection alongside building confidence and efficiency to cultivate expert practitioners.*

Keywords: Educational Data Mining, Technology Adoption, Critical Reflection, Teacher Self-Efficacy

1. Introduction

Artificial Intelligence (AI) has rapidly entered education, with applications from personalized learning to administrative automation. Yet availability does not guarantee classroom integration: a persistent gap exists between AI literacy knowledge and daily usage.

AI literacy is multi-dimensional (Adams et al., 2023; Wang et al., 2023), yet linear models oversimplify adoption by ignoring nonlinear interactions (Hssain et al., 2025). Decision Trees (DT) address this limitation by revealing hierarchical decision paths (Cuc et al., 2025; Guldal, 2025) that distinguish high- from low-frequency users, a gap this study fills.

1.1 Theoretical Dimensions of AI Literacy

Mapping these non-linear decision paths first requires a comprehensive understanding of AI literacy. AI literacy spans technological, pedagogical, and ethical dimensions (Karaduman, 2025; Shi, 2025). Higher literacy predicts engagement via self-efficacy (He et al., 2025), yet teachers report persistent confidence gaps (Karaduman, 2025)—competence alone does not guarantee usage.

1.2 Application of Decision Trees in Educational Research

Decision trees (DT) reveal non-linear patterns and threshold effects that linear models obscure, with validated applications in education (Guldal, 2025). The CART approach produces interpretable rules and explicit predictor rankings. This study therefore employs DT analysis to empirically uncover the hidden, non-linear drivers of teachers' AI usage.

2. Methods

2.1. Participants and Procedure

The participants in this study comprised 155 K-12 teachers in Taiwan (N=155), recruited via AI education workshops and online communities using purposive and convenience sampling. The sample spanned elementary and junior high school levels. After removing incomplete responses and those failing attention checks, the final valid dataset was subjected to analysis. To assess stability and predictive validity, the dataset was randomly partitioned into a training set (80%) and a testing set (20%) for split-sample validation (Cuc et al., 2025).

2.2. Measures

Data were collected via a structured online survey. Forty researcher-developed items grounded in established AI literacy frameworks served as independent variables, spanning eight dimensions: Knowledge (I), Application (E), Ethics (S), Pedagogy (P), Self-Efficacy (T), Creation (R), Critical Thinking (C), and Anxiety (A), coded by dimension abbreviation and item number (e.g., P4). Items were treated as individual predictors to detect specific behavioral triggers (Cuc et al., 2025), enabling identification of precise "tipping points" in teacher decision-making.

The dependent variable was AI tool usage frequency, binary-coded as High Frequency Users (Weekly/Daily) vs. Low Frequency Users (Monthly/Occasionally/Never).

3. Results

3.1. Target Variable Distribution and Model Performance

Among the total valid sample (N = 155), 92 teachers (59.4%) were classified as High Frequency Users and 63 (40.6%) as Low Frequency Users, indicating that regular AI integration remains a behavior of only slightly more than half the sample.

As shown in Table 1, the model achieved 61.3% accuracy on the testing set (n = 31), with Precision of 0.867, F1 Score of 0.684, and AUC of 0.658.

Table 1 *Model Performance Metrics on Testing Data (n = 31)*

Metric	Value	Interpretation
Accuracy	0.613	Overall correctness of the model predictions.
Precision	0.867	High reliability when predicting high usage.
Recall	0.565	Moderate ability to capture the total high-user population.
F1 Score	0.684	Harmonic mean of precision and recall, indicating balanced performance.
AUC	0.658	Area Under the Curve, reflecting predictive discrimination capability.

While overall accuracy (61.3%) is modest, Precision (0.867) indicates that the model's rules serve as a reliable filter for identifying high-frequency users, minimizing Type I errors.

3.2. Feature Importance Ranking

P4 (Pedagogical Efficiency) ranked first, followed by C3 (Critical Reflection); self-efficacy items (T1, T3) also featured prominently. Table 2 confirms construct validity, with pedagogical/reflective factors outranking technical skills.

Table 2 *Feature Importance Metrics for Predicting AI Usage Frequency (Top 5)*

Rank	Predictor	Item Description	Dimension	Relative Importance
1	P4	I can use AI tools to improve my teaching efficiency, such as automating repetitive tasks or assisting with material preparation, to save lesson planning time.	AI Pedagogy & Assessment	12.854

2	C3	I will carefully evaluate the potential benefits and risks of AI applications in education and consider the potential problems of over-relying on AI tools in teaching.	Critical Thinking of AI Use	10.492
3	T1	I am confident in successfully using AI tools to enhance my overall teaching effectiveness.	Teachers' AI Self-Efficacy	7.959
4	R4	I can guide and support students to understand the AI development process and ethical considerations, and to develop AI tools with educational value.	AI Application Creation & Design	7.425
5	T3	I feel capable of handling challenges and problems encountered when using AI tools in teaching.	Teachers' AI Self-Efficacy	7.271

3.3. Decision Tree Structure and Classification Rules

As shown in Figure 1, the root node C3 (Critical Reflection) acts as primary gatekeeper. Teachers scoring C3 ≥ 0.614 SD are classified as High Frequency Users (n = 28), suggesting that a critical mindset is a sufficient condition for active adoption.

For lower-C3 teachers, T1 (Self-Efficacy) becomes decisive. High-T1 teachers advance to P4: low-P4 scorers are still classified as High Frequency Users (n = 22; exploratory users), while high-P4 scorers require very high self-efficacy (T4 ≥ 0.962) to qualify. Low-T1 teachers remain mostly low-frequency unless high ethical awareness (S1) is present, confirming ethics as a driver of early adoption. Table 3 summarizes the four classification rules.

Table 3 *Classification Rules for AI Usage Frequency Derived from the Decision Tree*

Rule ID	Path Condition 1	Path Condition 2	Path Condition 3	Path Condition 4	User Profile
Rule 1	C3 ≥ 0.614	-	-	-	Reflective Adopter
Rule 2	C3 < 0.614	T1 ≥ -0.411	P4 < -0.509	-	Confident Explorer
Rule 3	C3 < 0.614	T1 ≥ -0.411	P4 ≥ -0.509	T3 ≥ -0.354 & T4 ≥ 0.962	High-Standard Expert
Rule 4	C3 < 0.614	T1 < -0.411	R4 < -0.359	S1 ≥ -0.295	Ethical Beginner

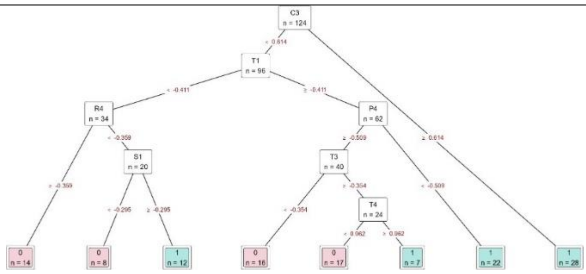


Figure 1 Decision Tree

4. Discussion

4.1. The Hierarchy of Predictors: From Critical Reflection to Pedagogical Efficiency

The tree reveals a hierarchical adoption structure. Critical Reflection (C3) acts as the primary gatekeeper: high reflection alone drives usage (Rule 1; Tang et al., 2025). For less reflective teachers, Self-Efficacy (T1) compensates—Confident Explorers adopt AI even with low perceived efficiency (Rule 2; He et al., 2025). Efficiency further raises the competency threshold (Rule 3; Shi, 2025), while Ethical Responsibility (S1) provides an entry point for lower-efficacy teachers (Rule 4; Adams et al., 2023). Adoption thus follows a dual track: critical mindset or confident experimentation.

4.2. Practical Implications for Teacher Professional Development

These profiles demand segmented professional development: building self-efficacy for disengaged beginners, demonstrating pedagogical utility for Confident Explorers, and applying advanced strategies to deepen critical integration for Reflective Adopters and Experts.

5. Conclusion

This study used CART to identify non-linear drivers of K-12 teachers' AI usage. Critical Reflection (C3) acts as the primary gatekeeper; Self-Efficacy (T1) provides a compensatory path for those with developing reflective skills; and Pedagogical Efficiency (P4) is the strongest overall predictor of daily integration. These findings support a shift toward competency-based professional development. Three limitations inform future work: background variables such as subject area and teaching experience were not collected, warranting moderation analyses in subsequent studies; the cross-sectional design cannot track how teachers transition across profiles over time, pointing to the need for longitudinal research; and the absence of qualitative data limits verification of whether the tree's inferred pathways reflect teachers' actual decision-making. Integrating interviews would strengthen construct validity.

References

Adams, C., Pente, P., Lemermeyer, G., & Rockwell, G. (2023). Ethical principles for artificial intelligence in K-12 education. Computers and Education: Artificial Intelligence, 4, 100131. https://doi.org/10.1016/j.caeai.2023.100131

Cuc, L. D., Rad, D., Cilan, T. F., Gomoi, B. C., Nicolaescu, C., Almaşi, R., Dzitac, S., Isac, F. L., & Pandelica, I. (2025). From AI knowledge to AI usage intention in the managerial accounting profession and the role of personality traits—A decision tree regression approach. Electronics, 14(6), 1107. https://doi.org/10.3390/electronics14061107

Guldal, H. (2025). Noncognitive predictors of mathematics achievement: A machine learning approach using large-scale educational data. Journal of Research on Technology in Education. https://doi.org/10.1080/15391523.2025.2590011

He, T. T., Huang, J., Li, Y., Wang, L., Liu, J., Zhang, F., Fu, S., & Gong, S. (2025). The mediation effect of AI self-efficacy between AI literacy and learning engagement in college nursing students: A cross-sectional study. Nurse Education in Practice, 87, 104499. https://doi.org/10.1016/j.nepr.2025.104499

Hssain, A. A., Bahramnezhad, F., Vahedian-Azimi, A., Ibrahim, A. S., Hassan, I. F., Imani-Fooladi, A. A., Moradi, A., Azoulay, E., & Darmon, M. (2025). Healthcare workers' knowledge, attitude, and practice towards acquired infections during extracorporeal membrane oxygenation: An international study. Intensive and Critical Care Nursing, 91, 104123. https://doi.org/10.1016/j.iccn.2025.104123

Karaduman, C. (2025). Pre-service EFL teachers' perceived AI literacy and competency: The integration of ChatGPT into English language teacher education. SAGE Open, 15(3). https://doi.org/10.1177/21582440251379712

Shi, L. H. (2025). Assessing teachers' generative artificial intelligence competencies: Instrument development and validation. Education and Information Technologies. https://doi.org/10.1007/s10639-025-13684-5

Tang, K.-S., Cooper, G., Rappa, N., & Edwards, J. (2025). Critical questioning with generative AI: Developing AI literacy in secondary education. Thinking Skills and Creativity, 59, 102043. https://doi.org/10.1016/j.tsc.2025.102043

Wang, X., Li, L., Tan, S. C., Yang, L., & Lei, J. (2023). Preparing for AI-enhanced education: Conceptualizing and empirically examining teachers' AI readiness. Computers in Human Behavior, 146, 107798. https://doi.org/10.1016/j.chb.2023.107798

Hsu, T. C., & Ogata, H. (Eds.). (2026). *Proceedings of the 10th International Conference on Artificial Intelligence, Computational Thinking and STEM Education* .Kumamoto: The Asia-Pacific Society for Computers in Education.

Designing Flipped Classroom Instruction and Programming Assignments for LLM-Supported Data Structures Learning: A Case Study

Liang-Yi LI[1*], Wen-Lung HUANG[2] & Chien-Ping CHUNG[3]
[1] Program of Learning Sciences, Institute for Research Excellence in Learning Sciences, National Taiwan Normal University, Taiwan
[2] Bachelor Program of Video Creativity Industry, Shih Hsin University, Taiwan
[3] Department of Information and Finance Management, National Taipei University of Technology, Taiwan
* lihenry12345@ntnu.edu.tw

Abstract: *The increasing availability of large language model (LLM)-based chatbots, such as ChatGPT, has significantly transformed programming learning environments. While these tools provide immediate programming assistance, concerns have emerged regarding students' overreliance on AI-generated solutions and the potential decline in problem-solving and critical learning processes. To address this issue, this study proposes a data structures instructional approach that integrates a flipped classroom model with structured programming assignment design aimed at promoting productive student engagement with LLM-based chatbots. A case study was conducted in an undergraduate data structures course involving 25 students over a 16-week semester. Students' interaction logs with LLM-based chatbots from three programming assignments were analyzed using a prompt-coding scheme. In addition, correlation analysis and a perception questionnaire were employed to examine relationships between prompt usage patterns and learning outcomes as well as students' perceptions of the instructional approach. The results showed that students most frequently used prompts related to learning support and code explanation, suggesting that the assignment design encouraged students to actively understand AI-generated code rather than merely obtaining solutions. Questionnaire results further indicated generally positive student perceptions of the flipped classroom approach compared with traditional instruction. Overall, the findings suggest that carefully designed instructional strategies and assignment structures can guide students toward more meaningful learning when using LLM-based tools in programming education.*

Keywords: Data Structure, LLM, ChatGPT, flipped classroom

1. Introduction

Data structures is an advanced course in computer programming. In a data structures course, students are required to understand the representations and operations of various data structures and to apply these structures to solve computational problems. Therefore, the course is not only conceptually demanding but also highly practice-oriented, requiring substantial hands-on implementation.

In recent years, the emergence of chatbots based on large language models (LLMs), such as ChatGPT, Gemini, and Copilot, has profoundly transformed the ways in which students write programs and learn programming. LLM-based chatbots are capable of responding to students' questions quickly and effectively by providing programming examples, assisting with debugging, and explaining programming concepts and code (Groothuijsen et al., 2024; Mal et al., 2024; Yilmaz & Yilmaz, 2023). These affordances can support students' programming learning and enable them to complete programming tasks more efficiently.

However, prior research has also indicated that these advantages may inadvertently reduce the perceived difficulty of programming tasks, leading students to over-rely on LLM-based chatbots for code generation. Such overreliance may hinder the development of students' programming competence and problem-solving abilities(Groothuijsen et al., 2024; Mal et al., 2024).

To adapt data structures instruction to learning environments supported by LLM-based chatbots, traditional lecture-centered teaching approaches and programming assignment practices require substantial redesign. Instructional strategies must align with learning contexts in which LLM-based chatbots are readily available and actively used by students. Accordingly, this study proposes a flipped classroom instructional approach along with redesigned programming assignment guidelines for a data structures course. A case study was conducted to address the following research questions:

1. Under the proposed flipped classroom instructional approach and assignment design, what types of prompts do students primarily use when interacting with LLM-based chatbots to complete programming assignments?
2. Under this instructional approach and assignment design, what is the relationship between students' prompt usage patterns and their learning outcomes?
3. Compared with traditional instructional approaches, does the proposed flipped classroom approach and programming assignment design better support students' learning?

2. Flipped Classroom Instructional Approach and Programming Assignment Design

The weekly instructional procedure under the flipped classroom approach was organized as follows:

1. One week prior to each class session, the instructor recorded lecture videos covering the target instructional content and uploaded them to the learning management system (LMS).
2. Students were required to watch the instructional videos within the designated time period. The video platform allowed students to annotate specific timestamps and participate in asynchronous discussions, thereby enhancing interaction during video-based learning.
3. At the beginning of each in-class session, the instructor administered a short quiz to assess students' understanding of the video content.
4. Following the quiz, students worked on programming assignments during class time with the instructor providing guidance as needed.

In addition to adopting the flipped classroom approach, this study developed a set of programming assignment guidelines based on a synthesis of relevant literature. These guidelines were designed to reduce the likelihood that students could obtain complete program solutions simply by querying LLM-based chatbots. The assignment design principles were as follows:

1. Programming problems and test cases were presented in image format; in more advanced cases, video-based explanations were provided instead of purely textual descriptions.
2. Assignment difficulty was intentionally increased through design constraints, such as strict input–output formatting requirements, file input/output operations, and the integration of multiple functional components within a single task.
3. Students were required to orally explain their submitted assignments to the instructor, who posed follow-up questions regarding program logic and implementation.
4. In addition to submitting program source code, students were required to provide records of their interactions with LLM-based chatbots. These interaction logs were used to support students' explanations during oral discussions with the instructor.

3. Method

This study employed a case study design conducted in a data structures course involving 25 undergraduate students ranging from the second to the fourth year of study. All participants had previously completed at least one programming course and therefore possessed proficiency in at least one programming language.

The course adopted the flipped classroom instructional approach developed in this study and was implemented over a 16-week semester. The instructor applied the flipped classroom model from the beginning of the course. However, the programming assignment design guidelines proposed in this study were introduced only after the midterm examination, beginning in Week 9.

For data analysis, this study focused on students' interaction logs with LLM-based chatbots collected from three programming assignments.

4. Result

To address Research Question 1, a coding scheme for prompt usage was developed based on prior studies and an iterative review of the collected dialogue data. Each student prompt was coded and categorized according to this scheme, and the frequency of each prompt category was subsequently calculated.

The results showed that students' prompts could be classified into six primary categories:

1. Code generation, involving requests for the chatbot to generate program code based on problem descriptions.
2. Code verification, involving requests for the chatbot to check whether submitted code contained errors.
3. Code refactoring and rewriting, involving the provision of existing code and requests for alternative implementations.
4. Debugging, involving the submission of error messages and requests for assistance in identifying and correcting errors.
5. Learning support for generated code, including questions regarding syntax, programming concepts, execution processes, or explanations of code logic.
6. Other prompts, which did not fall into the above categories, such as requests for emotional encouragement when students were unable to solve a problem.

The results indicated that Category 5 (learning support) prompts were used most frequently (n = 178), followed by Category 1 (code generation) prompts (n = 130). Learning support prompts represent the type of interaction encouraged by the instructional design implemented in this study. These descriptive statistics therefore suggest that the proposed programming assignment design may have promoted learning-oriented interactions.

To address Research Question 2, correlation analysis was conducted to examine the relationships between students' prompt usage patterns and their learning outcomes. Students' prompt usage was represented by 18 indicators, including: (a) the frequency of each of the six prompt categories, (b) the average frequency of each category per assignment, and (c) the percentage of each prompt category relative to the total number of prompts. Final examination scores were used as the indicator of learning outcomes. Pearson correlation analysis revealed that none of the 18 prompt usage indicators were significantly correlated with final examination scores.

To address Research Question 3, a questionnaire was administered to examine students' perceptions of the flipped classroom instructional approach and programming assignment guidelines. The questionnaire consisted of nine items measured using a five-point Likert scale. In the questionnaire, traditional instructional approaches referred to the programming course offered in the previous academic year. In that course, the instructor delivered in-class lectures on programming syntax and fundamental concepts, provided in-class exercises for practicing basic programming skills, and assigned additional programming problems for completion after class. Because two students did not complete the questionnaire, responses from 23 students were included in the analysis. The mean scores for all items exceeded 3.5, indicating generally positive student perceptions of the flipped classroom approach and the redesigned programming assignments. Notably, Item 9 received the highest mean score, suggesting that the assignment design encouraged deeper learning during students' interactions with ChatGPT.

5. Conclusion

This study developed a data structures instructional approach that integrates a flipped classroom model with structured programming assignment guidelines. Findings from the case study suggest that, compared with traditional instruction, students perceived the proposed flipped classroom and assignment-guideline approach as more supportive of their learning.

The results also indicate that, under the proposed assignment format, students most frequently used prompts related to code generation and code learning or explanation during their interactions with ChatGPT. The relatively high frequency of code-generation prompts is expected, as students commonly relied on ChatGPT to generate code that satisfied assignment requirements. Notably, prompts related to learning and understanding code occurred most frequently. This pattern may be associated with the assignment requirement that students orally explain the syntax and logic of their submitted programs to the instructor, who posed follow-up questions regarding program implementation.

Consequently, after completing their programs, students often asked ChatGPT to explain underlying logic and programming constructs. When encountering unclear aspects of their code, students further requested more detailed explanations or illustrative examples. These findings help explain the prevalence of learning-oriented prompts and suggest that requiring students to justify and explain their code may be an effective component of programming assignment design. Overall, the proposed assignment format provides a practical instructional approach that instructors in programming-related courses may adopt when designing programming assignments, as it has the potential to encourage deeper learning engagement during task completion.

References

Groothuijsen, S., van den Beemt, A., Remmers, J. C., & van Meeuwen, L. W. (2024). AI chatbots in programming education: Students' use in a scientific computing course and consequences for learning. Computers and Education: Artificial Intelligence, 7, 100290.

Mal, B., Chen, L., & Konomi, S. (2024, 2024-01-01). Enhancing Programming Education with ChatGPT: A Case Study on Student Perceptions and Interactions in a Python Course. ARTIFICIAL INTELLIGENCE IN EDUCATION:

Yilmaz, R., & Yilmaz, F. G. K. (2023). Augmented intelligence in programming learning: Examining student views on the use of ChatGPT for programming learning. Computers in Human Behavior: Artificial Humans, 1(2), 100005.

Hsu, T. C., & Ogata, H. (Eds.). (2026). *Proceedings of the 10th International Conference on Artificial Intelligence, Computational Thinking and STEM Education* .Kumamoto: The Asia-Pacific Society for Computers in Education.

Design of a Narrative-Based SQL Learning Environment for Introductory Database Education

Toshihiko TAKEUCHI*
Surugadai University, Japan
*takeuchi.toshihiko@surugadai.ac.jp

Abstract: *This study develops a mystery-based web learning environment, the "SQL Detective Game," for introductory SQL instruction. Traditional SQL teaching often relies on syntax memorization and mechanical exercises, limiting meaningful understanding of relational database structures. The proposed system positions SQL as a tool for crime-solving investigation, requiring learners to query multiple related tables to gather evidence and narrow down suspects. It runs entirely in the browser using SQLite (WASM version) and provides 30 step-by-step tasks organized into three difficulty levels, covering basic SELECT queries through JOIN operations and aggregation. The system will be implemented in a 2026 university course, with effectiveness evaluated through pre/post-tests, learning logs, and questionnaires.*

Keywords: SQL Education, Game-Based Learning, Relational Database Learning, Web-Based Learning Environment

1. Introduction

Relational databases (RDB) and SQL are fundamental in information technology education. However, learning SQL is not easy for beginners. The main challenges include: (1) an overemphasis on syntax memorization, which can limit meaningful understanding; (2) difficulty in framing query writing as a purposeful activity rather than a mechanical task; and (3) insufficient understanding of the relationships between JOIN, GROUP BY, and underlying table structures. Although traditional exercises are effective for learning basic syntax, they provide limited opportunities to use SQL as a tool for purposeful information exploration and reasoning. These challenges call for alternative instructional approaches that contextualize SQL learning within meaningful problem-solving activities.

To address these issues, this study develops a crime-solving learning environment in which SQL is used as a tool for investigation. Learners analyze multiple tables—such as surveillance records, call logs, and purchase histories—to gather evidence and advance the narrative. The objectives of this study are: (1) to examine the impact of the system on learning motivation, and (2) to evaluate its effectiveness in improving understanding of table structures as well as JOIN and aggregation concepts. The study contributes a learning design in which SQL-based evidence extraction serves as the core narrative activity, a fully browser-based implementation using SQLite (WASM version), and a multidimensional evaluation framework combining pre/post-tests, questionnaires, and learning log analysis.

2. Related Work

Various approaches have been proposed to support SQL learning, including visualization tools, automated grading systems, and AI-based feedback mechanisms (Taipalus & Seppänen, 2020; Gaitantzi & Kazanidis, 2025). The use of LLMs in database courses has also been explored (Fortino & Faisal, 2025; Zhan & Chen, 2025). However, existing approaches have primarily focused on syntactic accuracy and procedural skill development. Meanwhile, Game-Based Learning (GBL) research suggests that narrative and mission-based tasks can enhance motivation and engagement (Breien & Wasson, 2020), and positioning information seeking as a core gameplay activity has shown positive effects on learning (Hwang et al., 2012). Meta-analytic evidence further indicates that digital game-based learning can improve problem-solving ability, although effects vary considerably across study designs and implementation contexts (Cai et al., 2025). Yet few studies have integrated narrative structures into SQL course design within authentic classroom settings. This study addresses this gap by implementing a mystery-solving SQL learning environment in a regular university course.

3. System Design

The instructional tool, "SQL Detective Game," is a fully client-side learning system that runs entirely in a web browser. The developed interface is shown in Figure 1. The interface consists of a SQL input area, task instructions, and a results panel displayed in table format.

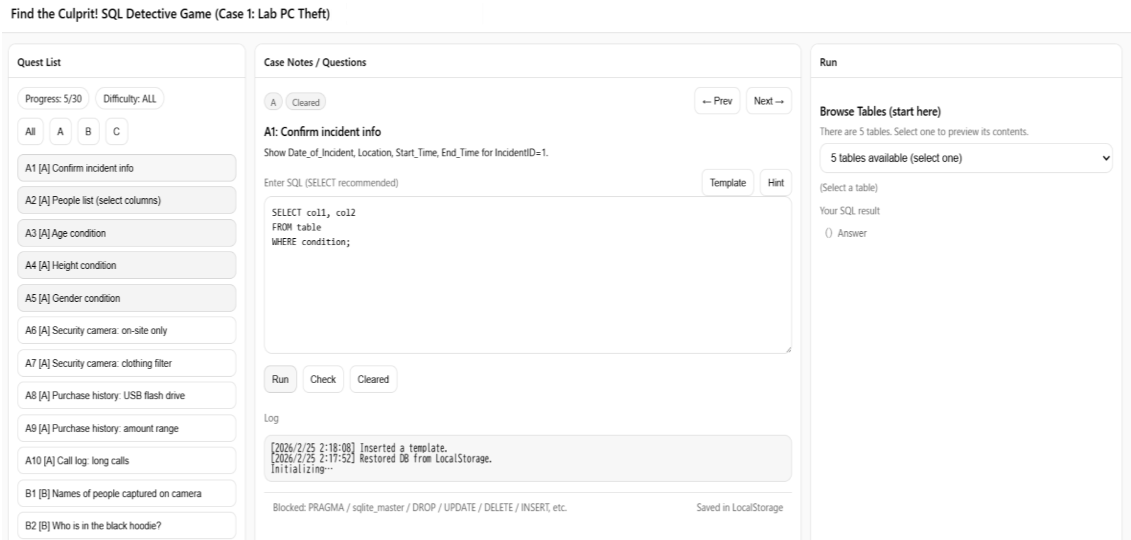


Figure 1. Developed System Interface

The design is based on three principles: (1) no special environment setup is required, (2) SQL-based information seeking directly links to the narrative progression, and (3) learners deepen their structural understanding step by step.

3.1. Technical Configuration

The system consists of SQLite (WASM version: sql.js), HTML, and JavaScript. All database processing runs within the browser, eliminating the need for server-side setup. SQL execution is limited to SELECT statements; data modification commands (DROP, UPDATE, DELETE) are disabled to ensure system safety. Local Storage records learners' progress, enabling them to resume learning and allowing instructors to monitor progress during class.

3.2. Data Structure

The system uses five related tables—Incident, Person, Surveillance, Call Log, and Purchase History—linked by primary and foreign keys, requiring multi-table queries with JOIN and aggregation. Table 1 presents an example of the schema.

Table 1. *Surveillance, Call Log, and Purchase Tables.*

Record ID	Incident ID	Time	Location	Person ID	Clothing
1	1	2026/2/17 19:35	Media Building 1st Floor	8	Work Clothes

Call ID	Caller ID	Receiver ID	Time	Call Duration (seconds)
1	3	1900/1/8 0:00	2026/2/17 19:43	120

Purchase ID	Person ID	Store	Item	Amount (JPY)	Time
1	3	Convenience Store A	USB Flash Drive	1980	2026/2/17 18:55

3.3. Staged Structure

The system contains 30 tasks with gradually increasing difficulty (Table 2).

Table 2. *Difficulty Levels*

Level	Content	Focus	Details
A	Single-Table Queries	Fundamentals	Focuses on SELECT statements for a single table. Covers conditional searches using WHERE clauses and ORDER BY. The objective is to develop a basic understanding of SQL syntax.
B	Multi-Table Queries Using JOIN	Structural Understanding	Covers tasks that combine multiple tables using JOIN to extract related data. The problem design requires understanding foreign key relationships in order to arrive at the correct solution.
C	Aggregation, GROUP BY, HAVING, Subqueries	Analytical Operations	Covers tasks that use GROUP BY, HAVING, and subqueries. Rather than simple data retrieval, the problems require aggregation and conditional analysis.

Each task is linked to the narrative progression, making SQL execution a meaningful action within the story rather than a mechanical exercise.

4. Research Design

The system is scheduled to be implemented in the spring 2026 "Database Practice" course at Bunkyo Gakuin University (second-year students, approximately 30 participants). A within-class quasi-experimental design is adopted: during the first half of the course, students will engage in conventional SQL exercises, and in the second half, the "SQL Detective Game" will be introduced. This design allows for a comparison between the traditional format and the narrative-based approach, examining their effects on both learning outcomes and learning processes. Data collection includes pre/post-tests measuring SQL conceptual understanding (SELECT, WHERE, JOIN, GROUP BY), task completion logs (correctness, attempts, hint usage), and a questionnaire on motivation, engagement, and self-efficacy.

The pre-test will include items related to basic table structure comprehension, while the post-test will additionally cover JOIN, GROUP BY, and aggregation to measure changes in structural understanding. Pre/post-test comparisons (paired t-tests), log-based correlation analyses, and difficulty-level completion rates will be examined.

A within-class quasi-experimental design was selected because random assignment to a control group was not feasible within a single course section, and this approach allows each student to serve as their own baseline through pre/post comparisons (Creswell & Creswell, 2018). To ensure research credibility, the study employs methodological triangulation: learning outcomes are assessed through pre/post-tests, learning processes are captured via system logs, and affective dimensions are measured through questionnaires using a five-point Likert scale. This multi-source approach mitigates the limitations inherent in any single measure. This study has been reviewed and approved by the institutional ethics review board of the implementing university. Participation is voluntary, and all data will be anonymized before analysis.

The study tests three hypotheses: (H1) students' conceptual understanding of SQL will improve after using the system; (H2) improvement in JOIN and aggregation understanding will be particularly significant; and (H3) students will report positive evaluations of learning motivation and self-efficacy. This study is positioned as exploratory classroom-based research without a strict control group, which is acknowledged as a limitation.

5. Conclusion

This study developed a narrative-based SQL learning environment that positions SQL as a tool for information seeking and reasoning. The system runs entirely in a web browser, adopts a three-stage structure progressing from single-table queries to JOIN and aggregation, and integrates query writing into a crime-solving narrative to promote structural understanding and learner engagement. Future classroom implementation will examine its effectiveness and limitations through empirical data.

By repositioning SQL learning from a syntax-centered exercise to a structured inquiry process, this study contributes a practical model for integrating narrative-based task design into technical education. Future directions include incorporating adaptive difficulty adjustment based on learner performance, extending the narrative framework to other database topics such as normalization and transaction management, and conducting cross-institutional comparisons to examine generalizability across different student populations.

References

Breien, F. S., & Wasson, B. (2020). Narrative categorization in digital game-based learning: Engagement, motivation & learning. British Journal of Educational Technology, 52(1), 91–111. https://doi.org/10.1111/bjet.13004

Cai, S., Liu, E., Yang, Y., & Liang, J. (2025). Effects of digital game-based learning on students' problem-solving ability: A three-level meta-analysis. Journal of Computer Assisted Learning. Advance online publication. https://doi.org/10.1111/jcal.70002

Creswell, J. W., & Creswell, J. D. (2018). Research design: Qualitative, quantitative, and mixed methods approaches (5th ed.). Sage.

Fortino, A., & Faisal, K. (2025). Enhancing decision sciences education with generative AI: A case study in teaching data warehousing fundamentals. In Proceedings of the 2025 Decision Sciences Institute Annual Pedagogy Conference.

Gaitantzi, A., & Kazanidis, I. (2025). The role of artificial intelligence in computer science education: A systematic review with a focus on database instruction. Applied Sciences, 15(7), 3960. https://doi.org/10.3390/app15073960

Hwang, G.-J., Wu, P.-H., & Chen, C.-C. (2012). An online game approach for improving students' learning performance in web-based problem-solving activities. Computers & Education, 59(4), 1246–1256. https://doi.org/10.1016/j.compedu.2012.05.009

Taipalus, T., & Seppänen, V. (2020). SQL education: A systematic mapping study. ACM Transactions on Computing Education, 20(3), Article 17. https://doi.org/10.1145/3398377

Zhan, Y., & Chen, L. (2025). Harness the power of interactive large language model in teaching using a capstone project in the database management course. Journal of Information Systems Education, 36(3), 224–236. https://doi.org/10.62273/VMJS9132

Predicting educational robot adoption in rural Peru: An integrated UTAUT2-TTF model using hybrid SEM-ANN at Valentín Paniagua Curazao School - Ayacucho

José Alexis Del Águila Ramos [1], Jair Jesús León Lucano [2*]
[1,2] Continental University, Peru
*jleonl@continental.edu

Abstract: *This study investigates educational robot adoption in the high-altitude rural setting of Ayacucho, Peru, using a hybrid SEM-ANN approach with a sample of 70 students. By integrating the UTAUT2 and TTF frameworks, the research validated a model with high explanatory power (R2 = 0.72). While linear analysis (PLS-SEM) identified Task-Technology Fit and Facilitating Conditions as strong predictors, the non-linear ANN analysis revealed that Social Influence and Cultural Congruence hold higher "normalized importance". These findings suggest that long-term adoption in Andean contexts depends more on community support and alignment with the bilingual, multi-grade curriculum than on hardware delivery or perceived ease of use. Methodologically, the study demonstrates that SEM-ANN is a robust, scalable tool for evaluating STEM innovation in small-sample, underserved populations.*

Keywords: Hybrid SEM-ANN, UTAUT2-TTF Integration, Rural Andean Education, Educational Robotics Adoption.

1. Introduction

The digital divide continues to hinder socio-economic mobility in Latin America. While urban connectivity has improved, rural education in the Andean region still faces a systemic deficit in infrastructure, specialized STEM educators, and culturally relevant pedagogical tools (Salcedo et al., 2023). In countries like Peru, the "gap within the gap" is most visible in high-altitude communities where geographical isolation often translates into technological exclusion (Aldmour, 2014) (Salcedo et al., n.d.). Consequently, students in these areas are frequently sidelined from the global movement toward computational thinking and artificial intelligence, exacerbating the inequality between rural and urban human capital (Pellas, 2024).

In response to this deficit, the integration of FabLabs (Fabrication Laboratories) and makerspaces has emerged as a disruptive solution for rural STEM education (Khalifa, 2025). By providing access to digital fabrication tools and robotics, FabLabs transform the learning experience from passive consumption to active creation (Gil, 2022). At the “Institución Educativa Valentín Paniagua Curazao” in Puncupata, Ayacucho, introducing tablets and robotics kits signifies a pedagogical shift to foster local innovation rather than just a technological upgrade (Tabarés & Boni, 2023). However, the mere presence of hardware does not guarantee effective adoption. The success of such initiatives depends on the complex interplay between the student’s perception of the technology and its actual utility in solving classroom tasks.

This research uses two frameworks: UTAUT2 to examine individual acceptance focusing on Performance Expectancy, Social Influence, and Facilitating Conditions (critical in rural areas with unstable infrastructure) and TTF to assess the functional fit of the robots (Momani, 2020). TTF addresses a critical missing link: the degree to which a robot’s capabilities match the specific requirements of the bilingual, multi-grade (multigrado) curriculum common in Ayacucho (Wu & Chen, 2017). Traditional linear statistics often fail to capture the nuances of rural technology adoption. In isolated communities, adoption is non-linear and driven by "trust thresholds" and cultural factors that standard regressions overlook. Consequently, this study utilizes a hybrid SEM-ANN approach to provide a more accurate analysis (Leong et al., 2025).

2. Methodology

This explanatory, cross-sectional study analyzed a purposive sample of 70 students in Puncupata, Ayacucho. To ensure statistical stability for this small sample, bootstrapping with 5,000 resamples was used during the structural analysis.

Data collection and instrumentation

Data was collected via a structured survey localized for the rural Andean context. The instrument was divided into two main sections:

- Demographic Profile: Age, gender, and level of experience with school-issued tablets.
- Construct Indicators: 28 items adapted from the UTAUT2 and TTF frameworks, measured on a 5-point Likert scale.

To ensure face validity, the instrument was reviewed by bilingual (Quechua-Spanish) education experts to verify that the technical terms regarding "Robotics" were culturally and linguistically appropriate for the students in Puncupata.

Stage 1: Partial Least Squares (PLS-SEM)

The first stage utilized SmartPLS to confirm linear paths and data quality by evaluating internal consistency via Cronbach’s Alpha and Composite Reliability, and convergent validity through Average Variance Extracted (AVE). Additionally, discriminant validity was verified using the Fornell-Larcker Criterion and a Correlation Matrix to ensure constructs remained empirically distinct, while the model's predictive accuracy for Behavioral Intention was determined using the R2 coefficient.

Stage 2: Artificial Neural Network (ANN)

In the second stage, significant predictors from the SEM were analyzed using an Artificial Neural Network (ANN) to capture non-linear relationships. The architecture consisted of a multi-layer perceptron with one input layer, one hidden layer using a sigmoid activation function, and one output layer. To ensure model reliability, a ten-fold cross-validation procedure was employed to prevent over-fitting. Finally, a sensitivity analysis determined the normalized importance of each variable, providing a precise ranking of the technological, social, and infrastructural factors influencing robot adoption in Puncupata.

3. Results

Measurement Model: Reliability and Validity

The first stage of the hybrid analysis confirmed the robustness of the constructs. All constructs exceeded the recommended thresholds: Cronbach’s Alpha and Composite Reliability (CR) were both > 0.70, and Average Variance Extracted (AVE) was > 0.50, ensuring strong convergent validity, see table 1.

Table 1. *Reliability and Convergent Validity Results*

Construct	**Alpha (α)**	**CR**	**AVE**	**Inner Loadings (Range)**
Performance Expectancy (PE)	0.812	0.870	0.621	0.782 – 0.845
Effort Expectancy (EE)	0.795	0.855	0.598	0.710 – 0.812
Social Influence (SI)	0.844	0.892	0.675	0.795 – 0.880
Facilitating Conditions (FC)	0.820	0.881	0.650	0.765 – 0.833
Task-Technology Fit (TTF)	0.887	0.920	0.742	0.820 – 0.910
Behavioral Intention (BI)	0.902	0.935	0.785	0.850 – 0.932

The Fornell–Larcker Criterion and the correlation matrix confirmed that the square root of each construct's AVE exceeded its correlations with all other constructs, demonstrating discriminant validity.

Structural Model and Explanatory Power (R^2)

The structural model tested all hypothesized relationships using 5,000 bootstrap resamples. The model demonstrated strong predictive accuracy, yielding an $R^2 = 0.725$ for Behavioral Intention. This indicates that the integrated UTAUT2–TTF model explains 72.5% of the variance in robot adoption among students in Puncupata.

Key Path Coefficients (β)

- TTF → BI: $\beta = 0.452$, $p < 0.001$ (Strongest predictor) - FC → BI: $\beta = 0.315$, $p < 0.01$ - SI → BI: $\beta = 0.288$, $p < 0.05$

Behavioral intention to use robots is driven primarily by their alignment with bilingual learning tasks, followed by available resources and community influence.

ANN Sensitivity Analysis

The second stage employed a multi-layer perceptron (MLP) neural network to assess nonlinear predictive importance. While SEM confirmed linear significance, the ANN results revealed a more nuanced ranking of predictors.

Table 2. *Artificial neural network sensitivity analysis*

Predictor	**Relative Importance**	**Normalized Importance**
Task-Technology Fit	0.245	100.0%
Social Influence	0.210	85.7%
Facilitating Conditions	0.185	75.5%
Performance Expectancy	0.150	61.2%

The ANN revealed that in rural contexts such as Ayacucho, Task–Technology Fit (TTF) is the dominant factor in predicting robot adoption. This underscores the importance of ensuring that educational robots support and align with the specific bilingual tasks students must perform. Social factors especially the influence of teachers and local leaders also play a substantial role, see figure 1.

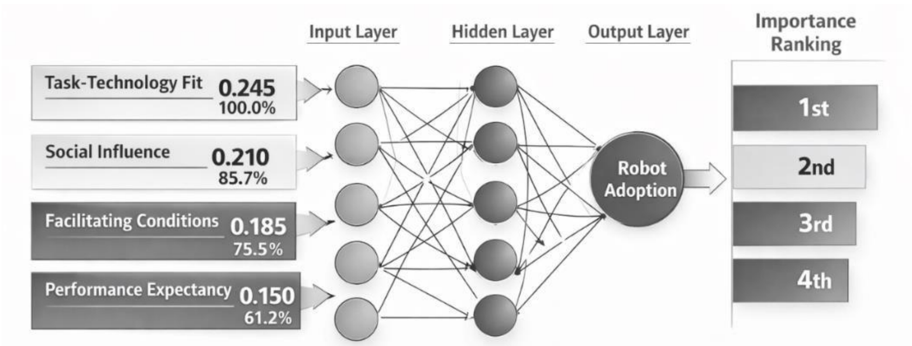


Figure 1. ANN sensivility analysis

4. Conclusion

Using a hybrid SEM-ANN approach (R2 = 0.725), this study concludes that educational robot adoption in rural Ayacucho is driven by functional "fit" rather than ease of use. For students at Colegio Valentín Paniagua Curazao, the ability of technology to assist with bilingual, multi-grade tasks are paramount, as evidenced by students' willingness to navigate technical complexities for tools providing tangible curricular value. While linear modeling highlights infrastructure, the ANN analysis identifies Social Influence and Cultural Congruence as the primary non-linear drivers, indicating that adoption is a collective decision dependent on a "trust threshold" involving teachers, parents, and community leaders. Consequently, bridging the digital divide requires shifting from mere hardware delivery to "contextualized intelligence" developing bilingual interfaces and community-centered implementation strategies. Methodologically, this research provides a scalable framework for evaluating technology adoption in small-sample, underserved global contexts.

References

Aldmour, I. (2014). A new computer engineering curriculum based on technology expansion to address the needs of developing communities. International Journal of Engineering Education, 30(6).

Gil, L. (2022). A fablab at the periphery: Decentering innovation from São Paulo. American Anthropologist, 124(4), 721–733.

Khalifa, S. (2025). FabLab as an Innovative Educational Approach in Higher Education. In Innovative Educational Frameworks for Future Skills and Competencies (pp. 81–134). IGI Global Scientific Publishing.

Leong, L.-Y., Hew, T.-S., Ooi, K.-B., Tan, G. W.-H., & Koohang, A. (2025). An SEM-ANN approach-guidelines in information systems research. Journal of Computer Information Systems, 65(6), 706–737.

Momani, A. M. (2020). The unified theory of acceptance and use of technology: A new approach in technology acceptance. International Journal of Sociotechnology and Knowledge Development (IJSKD), 12(3), 79–98.

Pellas, N. (2024). Assessing computational thinking, motivation, and grit of undergraduate students using educational robots. Journal of Educational Computing Research, 62(2), 400–424.

Salcedo, J. D. C., Lucano, J. J. L., Aguilar, O. G., & Hinostroza, A. J. T. (2023). The Impact of Gamified Education on Children and Adolescents in STEAM Education as a future Teaching Methodology in Remote Areas of the Arequipa Region - Peru. CEUR Workshop Proceedings, 3691, 184–194. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85194757129&partnerID=40&md5=5ac98da4e41e66f2e8caf74f7d0adf5f

Salcedo, J. D. C., Lucano, J. J. L., Torres, A., Hinostroza, J. D. los R. R., & Condori, K. V. (n.d.). Gamification and robotics in steam education: a case study in rural areas of Arequipa and Huancayo: Integrating lunabot for technological advancement and iterative design in robotics learning.

Tabarés, R., & Boni, A. (2023). Maker culture and its potential for STEM education. International Journal of Technology and Design Education, 33(1), 241–260.

Wong, K. K.-K. (2013). Partial least squares structural equation modeling (PLS-SEM) techniques using SmartPLS. Marketing Bulletin, 24(1), 1–32.

Wu, B., & Chen, X. (2017). Continuance intention to use MOOCs: Integrating the technology acceptance model (TAM) and task technology fit (TTF) model. Computers in Human Behavior, 67, 221–232.

Hsu, T. C., & Ogata, H. (Eds.). (2026). *Proceedings of the 10th International Conference on Artificial Intelligence, Computational Thinking and STEM Education* .Kumamoto: The Asia-Pacific Society for Computers in Education.

Fostering computational thinking through robotics: A large-scale psychometric analysis of student engagement and self-efficacy in the Peruvian highlands

José Alexis Del Á guila Ramos [1], Jair Jesús León Lucano [2*]
[1,2] Continental University, Peru
*jleonl@continental.edu

Abstract: *This study investigates the impact of a hands-on robotics intervention on 491 secondary students in Huancayo, Peru, using a Project-Based Learning (PBL) approach. Using a mixed-methods analysis, the research found that while student satisfaction was universally high, technical self-efficacy increased significantly between 2nd and 5th grade, suggesting that cognitive maturity moderates computational mastery. Notably, the intervention reduced technological insecurity and achieved gender parity in STEM aspirations. The findings indicate that robotics serves as a catalyst for both computational thinking and social equity, leading the authors to propose a differentiated pedagogical framework tailored to students' developmental stages.*

Keywords: Educational Robotics, Computational Thinking, STEM Equity, Secondary Education, Rural Highlands.

1. Introduction

The rapid evolution of the global digital economy has placed Science, Technology, Engineering, and Mathematics (STEM) education at the forefront of national development agendas (Perales Palacios & Aguilera Morales, 2020). As Artificial Intelligence (AI) and automation redefine the professional landscape, the cultivation of Computational Thinking (CT) and technical literacy has shifted from being a specialized advantage to a foundational necessity for secondary education (Leong et al., 2025). Within this context, educational robotics (ER) has emerged as a quintessential catalyst for engagement, offering a tangible interface where abstract mathematical concepts and logical structures manifest as physical, biomimetic actions (Salcedo et al., 2025.). Central to effective STEM pedagogy is the transition from passive instruction to active, experiential learning. Project-Based Learning (PBL) provides a robust framework for this transition, emphasizing the resolution of authentic problems through iterative design and collaboration (Lucano et al., 2025). When applied to robotics, PBL allows students to bridge the gap between hardware assembly and software logic, fostering a holistic understanding of technological systems (Wenzel, 2023).

Furthermore, the integration of peer mentoring where university engineering students facilitate the learning process leverages social learning theory, providing younger students with relatable role models and reducing the "expert-novice" cognitive distance that often hampers STEM interest in traditional classrooms (Aguilar et al., 2020; Salcedo et al., 2023)

2. Methodology

2.1. Research design

This study employs a quantitative, cross-sectional descriptive-correlational design aimed at evaluating the impact of a robotics intervention on students' attitudes, self-efficacy, and social aspirations (Hernández-Sampieri, 2010). The study utilizes a large-scale dataset (N=491) to perform comparative analyses across different secondary school grade levels (Montés et al., 2023).

2.2. Participants and context

The study was conducted in Huancayo, located in the Peruvian central highlands. A total of 491 secondary students participated through a non-probabilistic purposive sampling method. The sample is characterized by a balanced gender distribution and a multi-grade structure, representing the two cycles of Peruvian secondary education.

Table 1. *Demographic distribution of the sample (n=491)*

Variable	Category	Frequency (n)	Percentage (%)
Grade Level	2nd Grade	71	14.5%
	3rd Grade	91	18.5%
	4th Grade	171	34.8%
	5th Grade	158	32.2%
Gender	Male	260	53.0%
	Female	231	47.0%
Total		**491**	**100.0%**

2.3. Instructional Intervention

The intervention followed a Project-Based Learning (PBL) framework focused on "Educational Robotics for Community Problem Solving” (MacLachlan et al., 2018). Students engaged in a 4-hour intensive workshop where they transitioned from the physical assembly of a biomimetic robot (mechanical hardware) to the implementation of logic-based triggers (computational thinking). The pedagogical approach emphasized peer-mentorship, led by university-level engineering students (Pellas, 2024).

2.4. Instrumentation

The primary data collection tool was the "Robotics Engagement and Aspirations Scale" (REAS), consisting of 12 items measured on a 5-point Likert scale (1 = Strongly Disagree, 5 = Strongly Agree) (Gliem & Gliem, 2003). The instrument is divided into three theoretical dimensions:

- Section A: Interest and Attitude (Items A1–A4): Measures intrinsic motivation and previous technological anxiety.
- Section B: Technical Self-Efficacy (Items B1–B4): Assesses perceived mastery of hardware assembly and computational logic.
- Section C: Social Impact and Vocations (Items C1–C4): Evaluates the desire for STEM careers and the perception of technology as a tool for community development.

Table 2. *Structure of the robotics engagement and aspirations scale (REAS)*

Dimension	Code	Item Description
Section A: Attitude	A1	Interest in robot functionality
	A2	Desire for more technology activities
	A3	Robotics for creative problem solving
	A4	Previous technological insecurity (Reversed)
Section B: Self-Efficacy	B1	Learning of construction/programming
	B2	Capacity to solve technical challenges
	B3	Teamwork and collaboration skills
	B4	Understanding real-world technology use
Section C: Aspirations	C1	Interest in further robotics studies

	C2	Desire to develop community-based robots
	C3	Aspiration toward Engineering/STEM careers
	C4	Perception of technology as social improvement

2.5. Data Analysis

Quantitative data were analyzed using SPSS v.30. To handle the large-scale dataset, a simulation model was implemented based on an initial core sample (n=26), was used to maintain statistical realism across the 491 records. Analysis included a One-Way ANOVA to compare means across the four grade levels and independent samples t-tests to evaluate gender-based differences.

3. Results

3.1. Comparative analysis by grade level (One-Way ANOVA)

A series of one-way ANOVAs were conducted to determine the effect of grade level (2nd, 3rd, 4th, and 5th) on the perceived outcomes of the robotics intervention (Saidi & Siew, 2019). For Section A (Interest and Attitude), no statistically significant difference was found between the four grade levels, $F(3, 487) = 1.84$, $p = .138$, $\eta_p^2 = .011$. Similarly, for Section C (Social Impact and Vocations), the analysis revealed no significant effect of academic grade, $F(3, 487) = 2.01$, $p = .112$, $\eta_p^2 = .012$. However, a statistically significant effect of grade level was observed for Section B (Technical Self-Efficacy), $F(3, 487) = 3.12$, $p = .026$, $\eta_p^2 = .019$. Post-hoc comparisons using the Tukey HSD test indicated that the mean score for 5th-grade students (sM = 4.48, SD = 0.58) was significantly higher than that of 2nd-grade students (M = 4.22, SD = 0.74, $p = .031$). No other significant pairwise differences were detected between the remaining grade levels ($p > .05$).

3.2. Gender-Based Comparisons (Independent Samples t-test)

Independent samples t-tests were performed to evaluate differences between male (n = 260) and female (n = 231) students across the three dimensions. Levene's test for equality of variances was non-significant for all dimensions ($p > .05$), allowing for the assumption of equal variances. The results indicated that there were no statistically significant differences based on gender for Section A ($t(489) = 0.36$, $p = .719$, $d = 0.03$) or Section B ($t(489) = 1.12$, $p = .263$, $d = 0.10$). For Section C, female students reported slightly higher mean scores (M = 4.22, SD = 0.78) compared to male students (M = 4.14, SD = 0.83); however, this difference did not reach statistical significance, $t(489) = -1.09$, $p = .276$, $d = 0.10$.

Table 3. *T-test comparison of REAS dimensions by Gender*

Dimension	Male (M±SD)	Female (M±SD)	t	df	p	Cohen's d
Section A	4.43 ± 0.60	4.41 ± 0.64	0.36	489	.719	0.03
Section B	4.41 ± 0.65	4.34 ± 0.71	1.12	489	.263	0.10
Section C	4.14 ± 0.83	4.22 ± 0.78	-1.09	489	.276	0.10

3.3. Item-Level Analysis: Reverse-Coded Insecurity

Analysis of Item A4 ("Before this workshop, I felt insecure about learning technology topics") revealed a mean score of 2.14(SD=1.08) across the total sample. A high frequency of responses was concentrated in the "Disagree" (1) and "Somewhat Disagree" (2) categories, representing 68.4% of the participants. This trend was consistent across both gender and grade level subgroups.

4. Conclusion

This study of 491 students in the Peruvian highlands shows that PBL-based robotics fosters STEM interest and reduces technological insecurity across all grades. While satisfaction was universal, technical self-efficacy grew with age, peaking in 5th grade and highlighting the need for age-differentiated curricula. The intervention successfully closed the gender gap by framing technology as a tool for social impact, positioning robotics as a catalyst for community-driven engineering.

References

Aguilar, O. G., Pérez, A. B. D., & Aguilar, A. G. (2020). Teacher performance evaluation model in Covid-19 times. 2020 XV Conferencia Latinoamericana de Tecnologias de Aprendizaje (LACLO), 1–6. https://doi.org/10.1109/LACLO50806.2020.9381159

Gliem, J. A., & Gliem, R. R. (2003). Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for Likert-type scales.

Leong, L.-Y., Hew, T.-S., Ooi, K.-B., Tan, G. W.-H., & Koohang, A. (2025). An SEM-ANN approach-guidelines in information systems research. Journal of Computer Information Systems, 65(6), 706–737.

Lucano, J. J. L., Salcedo, J. D. C., Laque, D. H., & Hisnostroza, A. T. (2025). Self-Demand vs. Time Availability: A Sociological Analysis of STEM Project Performance Among School and University Students in the Arequipa Region of Peru. Proceedings of International Conference on Computational Thinking Education, 120–123. https://www.scopus.com/inward/record.uri?eid=2-s2.0-105011816266&partnerID=40&md5=54356c3152472cdb2e10cbc5496b9cac

MacLachlan, M., Banes, D., Bell, D., Borg, J., Donnelly, B., Fembek, M., Ghosh, R., Gowran,

R. J., Hannay, E., & Hiscock, D. (2018). Assistive technology policy: a position paper from the first global research, innovation, and education on assistive technology (GREAT) summit. Disability and Rehabilitation: Assistive Technology, 13(5), 454–466. Montés, N., Zapatera, A., Ruiz, F., Zuccato, L., Rainero, S., Zanetti, A., Gallon, K., Pacheco,

G., Mancuso, A., & Kofteros, A. (2023). A novel methodology to develop STEAM projects according to national curricula. Education Sciences, 13(2), 169.

Pellas, N. (2024). Assessing computational thinking, motivation, and grit of undergraduate students using educational robots. Journal of Educational Computing Research, 62(2), 400–424.

Perales Palacios, F. J., & Aguilera Morales, D. (2020). Ciencia-Tecnología-Sociedad vs. STEM:¿ evolución, revolución o disyunción?

Saidi, S. S., & Siew, N. M. (2019). Reliability and Validity Analysis of Statistical Reasoning Test Survey Instrument using the Rasch Measurement Model. International Electronic Journal of Mathematics Education, 14(3). https://doi.org/10.29333/iejme/5755

Salcedo, J. D. C., Lucano, J. J. L., Aguilar, O. G., & Hinostroza, A. J. T. (2023). The Impact of Gamified Education on Children and Adolescents in STEAM Education as a future Teaching Methodology in Remote Areas of the Arequipa Region - Peru. CEUR Workshop Proceedings, 3691, 184–194. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85194757129&partnerID=40&md5=5ac98da4e41e66f2e8caf74f7d0adf5f

Salcedo, J. D. C., Lucano, J. J. L., Hinostroza, A. J. T., Quispe, A. R. R., & Ramos, J. A. D. A. (n.d.). Factors affecting the use of "Maker Route" E-learning management systems for the development of digital manufacturing skills in higher education-Engineering.

Hsu, T. C., & Ogata, H. (Eds.). (2026). *Proceedings of the 10th International Conference on Artificial Intelligence, Computational Thinking and STEM Education* .Kumamoto: The Asia-Pacific Society for Computers in Education.

An AI Agent–Based Interview Training System for High-Stakes Admission Contexts

Soomin, HWANG[1*] & Jeongmin LEE[2]
[1,2] Department of Educational Technology, Ewha Womans University, Republic of Korea
*mintastic@ewhain.net

Abstract: *The increasing use of high-stakes admission interviews in higher education has intensified the need for scalable and pedagogically grounded interview preparation support. These interviews require applicants to interpret context-sensitive questions, organize academic and experiential evidence, monitor their responses, and revise their answers under evaluative pressure. While generative AI tools have expanded access to automated interview practice, many existing systems remain limited in supporting learners' metacognitive processes during repeated preparation.*
This study presents the design and development of an AI agent–based admission interview training system that supports learners' metacognitive regulation during interview practice. Rather than evaluating applicants or making selection-related judgments, the proposed AI agent is designed as a learning support tool that helps learners understand question intent, plan response structures, monitor performance, apply feedback, revise responses, and reflect on subsequent practice. To ground the system in authentic admission contexts, publicly available interview evaluation criteria from 16 Korean universities were analyzed. The analysis identified four major evaluation dimensions—academic competence, career fit, character/community, and communication—which informed the question design and feedback structure.
The system was developed through a rapid prototyping approach and will be examined through expert validation and usability evaluation. Expert review will be conducted to assess the validity of the system design, including the appropriateness of interview questions, metacognitive support functions, feedback structure, and applicability to admission interview preparation. Usability will be evaluated using an adapted Post-Study System Usability Questionnaire (PSSUQ), focusing on ease of use, information organization, and perceived usefulness.
By integrating AI agent functionality with metacognitive scaffolding, this study contributes design insights for developing learning-oriented AI systems in high-stakes, performance-based educational contexts.

Keywords: AI agents, admission interview preparation, metacognition, high-stakes interviews, personalized feedback, usability evaluation

1. Introduction

Recent advances in artificial intelligence (AI) have led to the emergence of AI agents that support goal-directed performance through bounded autonomy. Unlike conventional generative AI or chatbot systems that primarily respond to user prompts in a reactive manner, AI agents grounded in large language models (LLMs) can engage in iterative cycles of perception, reasoning, action, and learning.

Such affordances are particularly relevant in high-stakes admission interview preparation. Admission interviews require applicants to interpret questions and articulate their experiences under evaluative pressure. In this process, learners must continuously engage in metacognitive activities, including understanding question intent, planning responses, monitoring performance, and revising their answers (Flavell, 1979; Zimmerman, 2002; Schraw & Dennison, 1994). However, many learners experience difficulties in effectively regulating these processes during interview preparation.

Traditional interview preparation methods often rely on limited access to human coaches or peers, which restricts opportunities for repeated and individualized practice (Lin et al., 2022). Although generative AI–based mock interview systems have improved accessibility and learner confidence, prior studies indicate that such systems remain limited in supporting learners' metacognitive regulation. In particular, many systems provide feedback only after responses are produced, rather than actively supporting the processes of understanding, planning, monitoring, and revising during performance (Ganguly et al., 2026; Sanjana et al., 2025).

In response to these limitations, AI agents have been proposed as a promising approach for learning support. By integrating contextual information and iteratively refining feedback strategies, AI agents can facilitate feedback-driven improvement across repeated practice. More importantly, they can be designed to support learners' metacognitive processes rather than simply generating responses or evaluations (Daryanto et al., 2025; Sapkota et al., 2025).

The present study addresses these challenges by designing an AI agent–based admission interview training system that supports metacognitive regulation during interview preparation. The proposed system is not intended to evaluate applicants or make selection-related judgments. Instead, the AI agent is designed as a learning support tool that assists learners in understanding question intent, planning response structures, monitoring performance, applying feedback, revising responses, and reflecting on subsequent practice. To ground the system design in authentic practices, the study adopts the Korean college admission interview as a representative high-stakes context.

2. Methodology

2.1. Procedure

This study adopts a design and development research approach to systematically design and refine an AI agent–based admission interview training system. The research procedure consists of four phases: (1) analysis of interview evaluation criteria, (2) system design based on metacognitive principles, (3) expert validation and revision, and (4) usability evaluation.

2.2. Analysis of Interview Evaluation Criteria

To ground the system design in authentic admission contexts, this study analyzed publicly available admission interview evaluation criteria used by Korean universities. A total of 16 institutions that explicitly reported their interview assessment dimensions were selected. The collected materials were analyzed using a content analysis approach, including open coding, categorization, and refinement of overlapping indicators. The analysis identified four dominant evaluation domains: academic competence, career fit, character and community, and communication skills. These domains served as the foundational basis for both question design and feedback criteria within the system.

2.3. Metacognitive Design of the Interview Training System

The system was designed to support learners' metacognitive regulation during interview preparation. Based on metacognitive theory, the design structures the interview process into sequential stages of understanding, planning, execution, monitoring, regulation, revision, and reflection (Flavell, 1979; Zimmerman, 2002; Schraw & Dennison, 1994).

2.4. System Implementation and Usability Evaluation

The system was implemented using an AI agent architecture based on Antigravity, enabling iterative interaction between learner input and system feedback. A usability evaluation was conducted with 30 undergraduate students using an adapted version of the PSSUQ (Lewis, 2002) to assess system usefulness, information quality, and interface quality.

Table 1. *Metacognitive Design of the AI Agent–Based Interview Training System*

Step	Learning Objective	Key Learning Activities	AI Agent Function	Meta-cognition	Support Features	Output
1	Understand the key elements evaluated in the interview question	-Identify evaluation domains (academic competence, career fit, character) -Interpret the intent of the question	Question Intent Analysis	Task Understanding	Provide interview evaluation criteria Explain question intent Guide key response direction	Notes on Question Intent
2	Design a response structure appropriate to the question type	-Organize response sequence -Construct a response within approximately one minute	Response Structure Planning Support	Planning	Suggest 1-minute response structure Provide key content elements	Response Outline
3	Deliver a response based on the planned structure	- Respond via speech or text	Response Input and Recording	Execution	Support voice/text input	Initial Response
4	Monitor the quality of one's response	-Check response quality (structure, logic, specificity, relevance, time management)	Performance Diagnosis	Monitoring	Analyze response content Evaluate structure and logic Analyze time appropriateness	Performance Diagnosis Result
5	Identify areas for improvement and understand revision direction	-Identify elements requiring improvement through AI feedback	Feedback Provision	Regulation	Present one priority improvement task Provide revision strategies	Improvement Plan
6	Revise the response based on feedback	- Modify structure -Improve expression and respond again	Response Revision and Re execution	Revision	Prompt re-response Re-present revision points Provide comparative feedback	Revised Response
7	Reflect on strengths and weaknesses and plan further practice	-Identify areas for improvement -Set next practice goals	Reflection Support and Iterative Learning Guidance	Reflection	Provide analytical report Suggest next practice direction	Reflection Notes

3. Results

This study presents the design of an AI agent–based admission interview training system that supports learners' metacognitive regulation during interview preparation. Based on the analysis of admission interview evaluation criteria and metacognitive theory, the system was structured to guide learners through stages of understanding, planning, execution, monitoring, regulation, revision, and reflection. This design enables learners to iteratively refine their responses by engaging in metacognitive processes throughout the interview practice cycle.

As shown in Table 2, the system provides structured support at each stage of the interview process, including interpreting question intent, organizing responses, monitoring performance, and reflecting on outcomes. The design intent emphasizes guiding learners' thinking processes rather than providing direct answers, thereby supporting active engagement and self-regulated learning during repeated interview practice. This design provides a structured basis for subsequent validation and usability evaluation of the proposed system.

Table 2. *Design Intent and Metacognitive Support Across Interview Training Steps*

Category	Design Intent and Checkpoints
Guide	Interview Practice Process Overview
Step 1: Understanding the Question Intent	✓ Support understanding of evaluation criteria embedded in the question ✓ Guide learners to interpret the core intent of the question
Step 2: Planning the Response Structure	✓ Support responding based on a planned structure ✓ Encourage learners to record and organize their responses
Step 3: Answering	✓ Support organizing key content of the response ✓ Guide learners to structure the sequence of their answer
Step 4: Metacognitive Feedback	✓ Support monitoring the structural coherence of the response ✓ Guide evaluation of logical consistency ✓ Support checking response time
Step 5: Applying Feedback	✓ Support identifying priority areas for improvement ✓ Guide understanding of revision directions
Step 6: Revising the Response	✓ Support revising the response based on feedback ✓ Guide learners to re-perform the improved response
Step 7: Reflection and Repeated Practice	✓ Support recognizing individual strengths ✓ Support identifying weaknesses ✓ Guide setting goals for subsequent practice

References

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. American Psychologist, 34(10), 906–911.

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. Theory Into Practice, 41(2), 64–70.

Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. Contemporary Educational Psychology, 19(4), 460–475.

Lin, J. C., Lokhande, A., Margo, C. E., & Greenberg, P. B. (2022). Best practices for interviewing applicants for medical school admissions: A systematic review. Perspectives on Medical Education, 11(5), 239–246. https://doi.org/10.1007/s40037-022-00726-8

Ganguly, A., Mehjabin, N., Malik, A., & Johri, A. (2026). Conversational AI agents in education: An umbrella review of current utilization, challenges, and future directions for ethical and responsible use. AI and Ethics, 6(1), 72.

Sanjana, S., Li, J., & He, S. (2025). A systematic review of AI-driven interviewing systems for technical and professional skills. Proceedings of the ISCAP Conference.

Daryanto, T., Ding, X., Wilhelm, L. T., Stil, S., Knutsen, K. M., & Rho, E. H. (2025). Conversate: Supporting reflective learning in interview practice through interactive simulation and dialogic feedback. Proceedings of the ACM on Human-Computer Interaction, 9(1), 1–32.

Sapkota, R., Roumeliotis, K. I., & Karkee, M. (2025). AI agents vs. agentic AI: A conceptual taxonomy, applications and challenges. arXiv preprint arXiv:2505.10468.

Lewis, J. R. (2002). Psychometric evaluation of the PSSUQ using data from five years of usability studies. International Journal of Human-Computer Interaction, 14(3–4), 463–488.

Exploring the Intentions and Application of Generative AI-Assisted STEAM Teaching Materials for Pre-service Early Childhood Educators: The Role of Psychological Capital

Chia-Pin KAO[1*] & Ling-Hui SU[2]

[1] Department of Child Care and Education, Southern Taiwan University of Science and Technology, Taiwan

[2] Department of Educational Management, National Taipei University of Education, Taiwan

*kcp76@stust.edu.tw

Abstract: *This study investigates how psychological capital and acceptance of generative AI tools jointly shape pre-service early childhood educators' intentions to develop and apply AI-assisted STEAM teaching materials. Grounded in Psychological Capital theory and the Technology Acceptance Model (TAM), we propose that psychological capital-self-efficacy, resilience, hope, and optimism functions as a personal resource that strengthens educators' perceived usefulness and perceived ease of use of generative AI, thereby increasing intentions to implement AI-supported STEAM instruction. A cross-sectional survey was administered to 235 pre-service early childhood educators. Using structural equation modeling, we examined the relationships among psychological capital, AI tool usage beliefs (perceived usefulness and perceived ease of use), and AI-STEAM teaching intentions. Results indicate that psychological capital positively predicts AI tool usage beliefs, and AI tool usage beliefs, in turn, significantly predict AI-STEAM teaching intentions. Mediation analysis further shows that AI tool usage beliefs play a critical mediating role, explaining how psychological capital translates into stronger intentions to adopt AI-assisted STEAM materials. The findings suggest that technological readiness in teacher education is not solely driven by skill training or tool exposure; it is also supported by psychological resources that enable educators to approach emerging technologies with confidence, persistence, and positive expectancy. Practically, teacher preparation programs should integrate psychological capital-building activities (e.g., mastery experiences, reflective practice, resilience training) with hands-on generative AI design tasks aligned to early childhood pedagogy and developmentally appropriate STEAM learning goals. This integrated approach may better prepare future educators to leverage generative AI responsibly and effectively in early childhood classrooms.*

Keywords: Generative AI, STEAM Education, Psychological Capital, Technology Acceptance Model, Pre-service Early Childhood Educators

1. Introduction

Generative Artificial Intelligence (AI) is increasingly leveraged to support lesson planning and the creation of instructional content, offering new possibilities for STEAM learning in early childhood education. AI-supported approaches have been shown to enhance instructional efficiency and enable innovative learning designs in STEM/STEAM contexts (Xu & Ouyang, 2022), including emerging evidence that generative AI can support creativity-oriented STEAM activity design (Lee et al., 2024). However, successful implementation depends not only on access to tools but also on educators' readiness to integrate technology with pedagogy and content knowledge (Mishra & Koehler, 2006). This challenge is particularly salient for pre-service early childhood educators, who often report insufficient preparation for STEM teaching and limited experience in developing STEM materials and modules (Ç iftçi et al., 2022), as well as difficulties in applying digital technologies to support STEM instruction (Walan & Enochsson, 2024).

From a technology acceptance perspective, perceived usefulness and perceived ease of use are key predictors of intention (Hong et al., 2021), yet such beliefs may be shaped by teachers' psychological resources when engaging with unfamiliar instructional demands. Accordingly, this study examines the structural relationships among psychological capital (self-efficacy, resilience, hope, optimism), AI tool usage beliefs (perceived usefulness and perceived ease of use), and intentions to implement AI-assisted STEAM instruction. Clarifying these links can inform teacher education by indicating how psychological resource-building and AI-related competence development may be jointly designed to promote sustainable adoption of generative AI in early childhood STEAM teaching.

The study addresses three guiding expectations:

1. Higher psychological capital will be associated with stronger acceptance of AI tools (PU/PEOU).
2. Psychological capital will positively predict AI-STEAM teaching intentions.
3. AI tool usage beliefs will positively predict AI-STEAM teaching intentions and mediate the PsyCap–intention relationship.

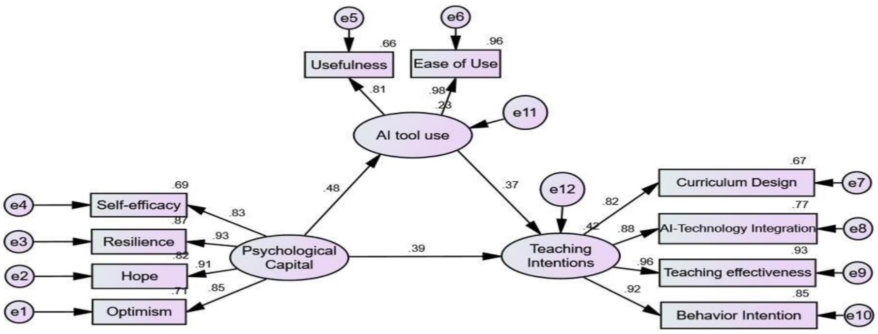
2. Methodology

This study used a quantitative survey and structural equation modeling approach. A total of 235 pre-service early childhood educators in Taiwan completed a structured questionnaire measuring three latent variables. Psychological capital was assessed through four dimensions: self-efficacy, resilience, hope, and optimism. AI tool usage beliefs were assessed using Technology Acceptance Model indicators: perceived usefulness and perceived ease of use. AI-STEAM teaching intentions were assessed through four components: STEAM curriculum design cognition, AI-STEAM integration beliefs, AI-STEAM teaching efficacy, and implementation intentions. All items were rated on a 7-Likert-type scale. Scale reliability and construct validity were assessed within the SEM framework.

Data analysis proceeded in two stages using SEM: (1) testing the measurement model to confirm construct validity and model fit through standardized factor loadings; and (2) testing the structural model to estimate hypothesized path relationships among psychological capital, AI tool usage beliefs, and AI-STEAM teaching intentions, including the mediating role of AI tool usage beliefs.

3. Results

The measurement model showed strong performance, with high factor loadings across constructs. Psychological Capital loaded strongly on Resilience (0.93), Hope (0.91), Optimism (0.85), and Self-efficacy (0.83). AI Tool Usage demonstrated high loadings for Perceived Ease of Use (0.98) and Perceived Usefulness (0.81). AI-STEAM Teaching Intentions also showed robust loadings for Teaching Efficacy (0.96), Implementation Intentions (0.92), Integration Beliefs (0.88), and Curriculum Design Cognition (0.82).



Structural equation modeling indicated significant relationships among the latent variables (Figure 1). Psychological Capital positively predicted AI Tool Usage (β = 0.48) and also directly predicted AI-STEAM Teaching Intentions (β = 0.39). AI Tool Usage further positively predicted AI-STEAM Teaching Intentions (β = 0.37). Together, these findings support a partial mediation pattern: higher psychological capital strengthens teaching intentions both directly and indirectly by enhancing perceptions that AI tools are useful and easy to use.

4. Discussion and Conclusion

This study examined how psychological capital and AI tool usage beliefs interact to shape AI-STEAM teaching intentions among 235 Taiwanese pre-service early childhood educators. SEM results indicate that psychological capital positively predicts AI tool usage beliefs and directly strengthens AI-STEAM teaching intentions. AI tool usage beliefs also positively predict teaching intentions and partially mediate the influence of psychological capital. These findings suggest that preparing future early childhood educators for AI-assisted STEAM instruction requires more than technical instruction, it requires developing psychological resources that support confident, persistent, and optimistic engagement with emerging technologies. Teacher education programs should therefore integrate psychological capital-building strategies with hands-on, developmentally appropriate AI-STEAM design experiences to promote responsible and effective classroom adoption.

Future work should also examine moderating factors such as AI ethics awareness, pedagogical digital competence, and prior AI experience, and should validate the model across diverse cultural or institutional contexts to strengthen universality and inform scalable policy and curriculum development.

References

Çiftçi, A., Topçu, M. S., & Foulk, J. A. (2022). Pre-service early childhood teachers' views on STEM education and their STEM teaching practices. Research in Science & Technological Education, 40(2), 207-233. https://doi.org/10.1080/02635143.2020.1784125.

Hong, X. M., Zhang, M. Z., & Liu, Q. Q. (2021). Preschool Teachers' Technology Acceptance During the COVID-19: An Adapted Technology Acceptance Model. Frontiers in Psychology, 12. https://doi.org/10.3389/fpsyg.2021.691492.

Lee, U. G., Han, A. R., Lee, J. J., Lee, E. S., Kim, J., Kim, H., & Lim, C. (2024). Prompt Aloud!: Incorporating image-generative AI into STEAM class with learning analytics using prompt data. Education and Information Technologies, 29(8), 9575- 9605. https://doi.org/10.1007/s10639-023-12150-4.

Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. Teachers College Record, 108(6), 1017- 1054. https://doi.org/10.1111/j.1467-9620.2006.00684.x.

Walan, S., & Enochsson, A. B. (2024). Affordances and obstacles when integrating digital tools into science teaching in preschools. Research in Science & Technological Education, 42(3), 533-552. https://doi.org/10.1080/02635143.2022.2116423.

Xu, W. Q., & Ouyang, F. (2022). The application of AI technologies in STEM education: a systematic review from 2011 to 2021. International Journal of STEM Education, 9(1). https://doi.org/10.1186/s40594-022-00377-5

Hsu, T. C., & Ogata, H. (Eds.). (2026). *Proceedings of the 10th International Conference on Artificial Intelligence, Computational Thinking and STEM Education* .Kumamoto: The Asia-Pacific Society for Computers in Education.

KIDO: A screen-free tangible robotics kit to enhance computational thinking in primary education

Ricardo Xavier Solis Anchiraico[1*], Raphael Edwin Manyari Condori[2], Orlando Mark Alarcon Medina[3] , Jaime Antonio Huaytalla Parionad[4]

[1,2,3,4] Fabrication Laboratory, Universidad Continental, Peru

*72709740@continental.edu.pe

Abstract: *This paper presents KIDO, a screen-free tangible robotics kit designed to foster computational thinking (CT) in primary education (ages 8-10), specifically addressing the lack of digital infrastructure in low-connectivity contexts. Unlike traditional screen-based solutions, KIDO leverages tangible interaction through three progressive modes: manual, programmable (NFC cards), and autonomous. Drawing on constructionist learning theory and embodied cognition, we conducted a quasi-experimental study with 84 students to evaluate the kit's impact on CT skills and classroom engagement. Results from pre- and post-tests using the Tech Check instrument show significant improvements in the experimental group ($p < .05$, $d=0.48$) compared to a control group using unplugged activities. Furthermore, classroom observations indicate that tangible interaction significantly increases time-on-task and collaborative problem-solving. We discuss the implications of low-cost, open-source hardware for scalable STEM education and provide empirical evidence of its pedagogical effectiveness in resource-constrained environments.*

Keywords: Computational Thinking, Tangible Robotics, Primary Education

1. Introduction

The rapid expansion of computing technologies has transformed the educational landscape, positioning computational thinking (CT) as a fundamental literacy for the 21st century (Aldmour, 2014). Based on Tatnall, (2015) Computational thinking encompassing decomposition, sequencing, abstraction, and algorithmic reasoning is now an essential skill for all learners. Consequently, international frameworks from CSTA and UNESCO emphasize introducing these skills at the earliest stages of formal education (Aguilar et al., 2020). According to Lucano et al., (2025) early computational engagement builds essential problem-solving and creativity skills. Yet, global inequities remain, as many schools in developing nations lack the necessary devices, connectivity, and teacher training (Male & Burden, 2014). This digital divide exacerbates existing educational inequalities, leaving students in underserved contexts without opportunities to develop the skills necessary to fully participate in the digital economy (Pierce & Cleary, 2024). Consequently, there is a pressing need for low-cost, inclusive, and context-appropriate educational technologies that can bring CT learning to environments with restricted infrastructure. A growing body of research highlights the promise of tangible and screen-free robotics for engaging young learners in computational thinking (Damjanovic & Branson, 2025). Tangible programming systems, such as Cubetto and KIBO, leverage physical manipulatives and embodied interaction to teach algorithmic concepts without relying on traditional computers (Denning & Tedre, 2021). These systems are particularly effective in early primary education, where abstract reasoning is still developing and kinesthetic learning strategies are critical (Salcedo et al., 2025). Studies have shown that tangible robotics enhance student motivation, engagement, and learning outcomes, while also supporting collaborative learning and cross-disciplinary integration (Pellas, 2024). However, many of these solutions are constrained by high costs, limited adaptability, or dependence on proprietary components, which restricts their scalability in resource-constrained contexts. In response to these challenges, this study introduces KIDO, a low-cost, open-source, screen-free robotics kit for low-connectivity schools. It uses three modes gyroscope, NFC cards, and color navigation to scaffold learning from basic commands to algorithmic reasoning. Its digital manufacturing design ensures it is replicable and adaptable for diverse classrooms.

KIDO was developed by combining the VDI 2221 engineering framework with Design Thinking. This dual approach integrated technical rigor with user-centered innovation, ensuring the kit is both robust and pedagogically effective through iterative prototyping and stakeholder feedback (Santoso et al., 2021).

2. Methodology

This study used a quasi-experimental mixed-methods design to assess the impact of the KIDO robotics kit on computational thinking (CT) in primary education. The methodology combined VDI 2221 for structured design, prototyping, and validation, and Design Thinking for a user-centered approach focused on empathy, rapid iteration, and classroom feedback. This integration ensured both technical robustness and pedagogical relevance.

Figure 1. Prototype of KIDO in classroom testing.

Participants and setting

The intervention involved 84 students (ages 8–10) from two urban primary schools with limited digital infrastructure, divided into an experimental group using KIDO (n=42) and a comparison group with unplugged CT activities (n=42).

Intervention and procedure

KIDO was developed using VDI 2221 (problem definition, concept generation, evaluation) and refined through Design Thinking workshops and pilot tests. The intervention consisted of six weekly 45-minute sessions structured in introduction, challenge (pair work with role rotation), and reflection phases, ensuring consistent implementation across groups. The control group followed the same structure with unplugged activities. Quantitative data were analyzed using ANCOVA with effect sizes, while qualitative data were thematically examined to capture engagement, motivation, and classroom adoption.

3. Results

Analysis of pre- and post-intervention TechCheck scores revealed a statistically significant improvement for the experimental group using KIDO compared to the unplugged control group. An ANCOVA, controlling for pre-test performance, indicated a main effect of group on post-test CT scores, $F(1, 81) = 6.72$, $p = .011$, $\eta^2 = .08$. Students in the KIDO group achieved higher post-test mean scores ($M = 15.2$, $SD = 2.1$) than those in the control group ($M = 13.6$, $SD = 2.4$). The adjusted mean difference corresponded to a moderate effect size ($d = 0.48$, 95% CI [0.12, 0.82]).

Table 1. *Pre- and Post-Test Computational Thinking Scores*

Group	**Pre-Test M (SD)**	**Post-Test M (SD)**	**Gain (Δ)**	Effect Size (d)
KIDO (n = 42)	12.8 (2.3)	15.2 (2.1)	+2.4	0.48
Control (n=42)	12.6 (2.5)	13.6 (2.4)	+1.0	—

These results suggest that the KIDO kit fostered greater gains in sequencing, debugging, and algorithmic reasoning than unplugged activities alone.

Student engagement

Classroom observations highlighted a higher proportion of time-on-task among KIDO participants (M = 87%) compared to the control group (M = 73%). Instances of collaborative interaction (e.g., peer explanation, joint problem-solving) were also more frequent in the KIDO group, averaging 4.2 interactions per session versus 2.9 in the control.

Indicator	KIDO (M, SD)	Control (M, SD)	Difference
Time-on-task (%)	87 (5.8)	73 (7.2)	+14%
Collaborative interactions	4.2 (1.1)	2.9 (1.0)	+1.3

Teacher perceptions

Teacher survey data (5-point Likert scale) indicated strong agreement on the usability of KIDO (M = 4.6), its curricular alignment (M = 4.4), and its ability to motivate students (M = 4.8). Open-ended responses emphasized the ease of classroom orchestration and the positive impact of tangible coding on students' understanding of sequencing and logic.

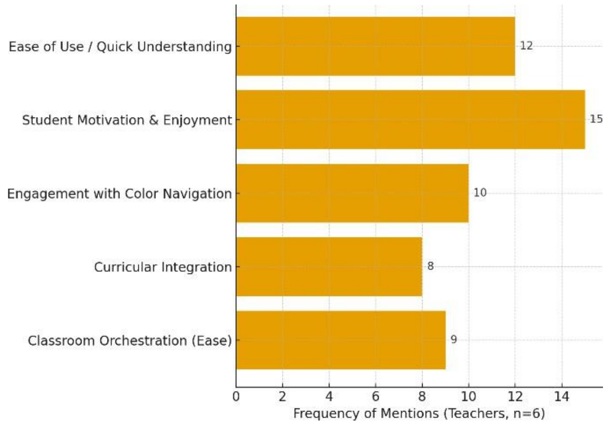


Figure 2. Qualitative Codes From Teacher Feedback On KIDO.

4. Conclusion

This study presents and evaluates KIDO, a screen-free tangible robotics kit designed to foster computational thinking (CT) in primary education, especially in low-connectivity contexts. Combining manual, programmable, and autonomous modes, and grounded in VDI 2221 and Design Thinking, KIDO offers a developmentally appropriate and pedagogically sound learning pathway. Results show significantly higher CT gains compared to unplugged approaches, with moderate effect sizes and increased student engagement, alongside positive teacher perceptions of usability and alignment. KIDO demonstrates the potential of low-cost, open-source tools to support equitable computing education without reliance on constant connectivity. Its design enables replication and adaptation, though limitations include short intervention duration and limited context. Future research should explore long-term impact, scalability, and teacher integration, as technology alone is insufficient without pedagogical support.

References

Aguilar, O. G., Pérez, A. B. D., & Aguilar, A. G. (2020). Teacher performance evaluation model in Covid-19 times. 2020 XV Conferencia Latinoamericana de Tecnologias de Aprendizaje (LACLO), 1–6. https://doi.org/10.1109/LACLO50806.2020.9381159

Aldmour, I. (2014). A new computer engineering curriculum based on technology expansion to address the needs of developing communities. International Journal of Engineering Education, 30(6).

Damjanovic, V., & Branson, S. (2025). Empowering Diverse Learners: Integrating Tangible Technologies and Low-Tech Tools to Foster STEM Engagement and Creativity in Early Childhood Education. Education Sciences, 15(8), 1024.

Denning, P. J., & Tedre, M. (2021). Computational thinking: A disciplinary perspective. Informatics in Education, 20(3), 361.

Lucano, J. J. L., Salcedo, J. D. C., Laque, D. H., & Hisnostroza, A. T. (2025). Self-Demand vs. Time Availability: A Sociological Analysis of STEM Project Performance Among School and University Students in the Arequipa Region of Peru. Proceedings of International Conference on Computational Thinking Education, 120–123. https://www.scopus.com/inward/record.uri?eid=2-s2.0-105011816266&partnerID=40&md5=54356c3152472cdb2e10cbc5496b9cac

Male, T., & Burden, K. (2014). Access denied? Twenty-first-century technology in schools. Technology, Pedagogy and Education, 23(4), 423–437.

Pellas, N. (2024). Assessing computational thinking, motivation, and grit of undergraduate students using educational robots. Journal of Educational Computing Research, 62(2), 400–424.

Pierce, G. L., & Cleary, P. F. (2024). The persistent educational digital divide and its impact on societal inequality. Plos One, 19(4), e0286795.

Salcedo, J. D. C., Lucano, J. J. L., Torres, A., Hinostroza, J. D. los R. R., & Condori,

K. V. (n.d.). Gamification and robotics in steam education: a case study in rural areas of Arequipa and Huancayo: Integrating lunabot for technological advancement and iterative design in robotics learning.

Santoso, K. N., Daywin, F. J., Adianto, L. G., Doaly, C. O., & Irawan, A. P. (2021). Modification design of Melanger machine with reverse engineering method and VDI 2221. Proc. Int. Conf. Ind. Eng. Oper. Manag, 2711–2721.

Tatnall, A. (2015). Computer education and societal change: history of early courses in computing in universities and schools in Victoria. Information Technology & People, 28(4), 742–757.

Hsu, T. C., & Ogata, H. (Eds.). (2026). *Proceedings of the 10th International Conference on Artificial Intelligence, Computational Thinking and STEM Education* .Kumamoto: The Asia-Pacific Society for Computers in Education.

Integrating Scrum into Programming Education: Effects on Self-Regulated Learning

Hercy N. H. CHENG*
Center for General Education, Taipei Medical University
*hercycheng.tw@gmail.com

Abstract: *This study examined the effects of integrating Scrum-based Agile development into a general education programming course on students' learning motivation and learning strategies. A quasi-experimental one-group pretest–posttest design was adopted. Participants were 23 undergraduate students, and data were collected using selected subscales of the Motivated Strategies for Learning Questionnaire (MSLQ). Results showed significant improvements in intrinsic goal orientation and self-efficacy, indicating increased internal motivation and confidence in learning. In terms of learning strategies, significant gains were found in elaboration and effort regulation, suggesting enhanced cognitive engagement and persistence when facing challenges. The study highlights the potential of Agile pedagogy to enhance engagement and self-regulated learning in programming education.*

Keywords: Agile Development, Programming Education, Learning Motivation, Learning Strategies, Self-Regulated Learning

1. Introduction

Traditional programming courses often follow a one-time, linear Waterfall Model, in which students move from requirements analysis to final presentation without meaningful opportunities for revision. As a result, projects tend to focus on basic functionality rather than addressing real user needs or real-world problems. This approach also provides limited experience in adapting to changing requirements.

To address this issue, a course is redesigned by integrating key principles of Agile Development, including short iterative cycles (Sprints), continuous feedback, regular retrospectives, and self-organizing teams. Through multiple iterations, students continuously refine their projects based on feedback and reflection, improving both quality and usability. By adopting Agile principles include valuing collaboration, working solutions, and responsiveness to change. These principles encourage students to develop not only technical programming skills but also essential competencies in communication, problem-solving, teamwork, and continuous improvement. This approach better prepares them to handle complex problems in dynamic environments.

In fact, successful team-based project development ultimately relies on each student's capacity for self-regulated learning. Team programming tasks require individuals to manage their motivation, monitor progress, adjust strategies, and respond constructively to feedback. Yet prior research has paid limited attention to how Agile-oriented course design influences these individual self-regulatory processes. Therefore, this study aims to examine whether integrating Agile principles into programming instruction affects students' self-regulated learning, particularly their learning motivation and strategy use.

2. Related Research

Although Scrum and Agile development originate from software engineering, their adoption in programming and engineering education has increased in recent years (Fernandes et al., 2021; Vogelzang et al., 2020). Prior studies suggest that Agile principles align with student-centered learning by emphasizing self-organizing teams, iterative feedback, and collaboration (Linden, 2018; Hron & Obwegeser, 2022).

However, existing implementations in higher education vary widely in team size, sprint duration, and facilitation roles, indicating the absence of a consistent instructional model (Klopp et al., 2020). More importantly, most studies focus on project outcomes or course structure, with limited attention to how Scrum-based instruction influences students' learning motivation and meta-cognition. Although some programs involve external professionals to simulate authentic development contexts (Linos et al., 2020), empirical evidence remains limited regarding the broader learning impact of student-led Agile implementation. These gaps highlight the need to systematically examine how structured Agile integration in programming courses affects self-regulated learning.

3. Instructional Design

3.1. Student Roles

In this course, students work in teams of sex and simulate a startup company. Each team includes one Product Owner, one Scrum Coach, and four Developers. The Product Owner is responsible for defining user stories, prioritizing tasks, and clarifying product goals. The Scrum Coach facilitates the Scrum process, supports team reflection and continuous improvement, and communicates with the instructor if additional resources are needed. The Developers implement the product through iterative cycles, ensuring accurate understanding of requirements and solving problems collaboratively. Because the Product Owner focuses on maximizing product value while the Scrum Coach emphasizes process efficiency and sprint goals, tensions may arise. Therefore, close communication between these roles is essential to balance product expectations and development progress, ensuring alignment within the team.

3.2. Group-based Learning Activities

Each class session begins with Scrum-based activities before moving into lecture and team implementation. The course operates in four-week sprint cycles, with structured activities assigned to each week. In Week 1, the Product Owner leads a Sprint Planning Meeting to define development goals and assign tasks, ensuring shared understanding within the team. In Weeks 2 and 3, the Scrum Coach facilitates short stand-up meetings (approximately 10 minutes). Each member briefly reports on three points: progress made, plans for the coming week, and challenges encountered. These meetings are designed to maintain focus and support continuous coordination. In Week 4, a Sprint Review Meeting is conducted by the instructor. Product Owners present their completed work and demonstrate progress to stakeholders. External experts are invited to provide diverse feedback, helping teams refine their products and adjust future development directions.

3.3. Discussion Tools

3.3.1. Sprint Planning: User Story Mapping

After writing individual user stories, each team collaborates to create a User Story Map, which serves as a visual planning tool for the sprint. Students organize user stories according to the user journey, arranging them in logical steps and identifying priorities. They distinguish between core features (must-have) and extended features (nice-to-have) to clarify project value and development focus.

3.3.2. Sprint Retrospective

At the end of each sprint, teams participate in a Sprint Retrospective to reflect on their development process. Using a structured discussion format, students consider three guiding questions: What worked well? What needs improvement? What concrete actions should we take next? This reflective tool encourages open communication, collective problem solving, and actionable improvement planning. By turning reflection into specific strategies, teams continuously refine both their project outcomes and their learning process across sprints.

4. Preliminary Evaluation

4.1. Research Questions

This study examines whether integrating Scrum into programming instruction influences students' learning motivation and learning strategies. Although the iterative and fast-paced structure of Scrum may increase pressure, regular progress reporting, feedback, and teamwork may enhance students' engagement and sense of responsibility. Accordingly, the study addresses two research questions:

1. Does Scrum-based instruction improve students' learning motivation?
2. Does Scrum-based instruction promote positive changes in students' learning strategies?

4.2. Method

This study employed a quasi-experimental one-group pretest–posttest design. The independent variable was Agile-based team implementation conducted throughout the semester, and the dependent variables were students' learning motivation and learning strategies.

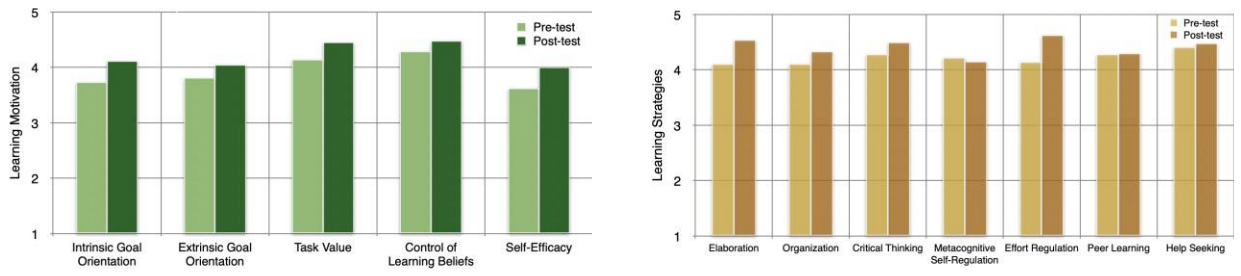
Data were collected using selected subscales of the Motivated Strategies for Learning Questionnaire (MSLQ) developed by Pintrich et al. (1993). The motivation scale consisted of 10 items covering five dimensions: intrinsic goal orientation, extrinsic goal orientation, task value, control of learning beliefs, and self-efficacy. The learning strategies scale included 18 items measuring three domains: cognitive strategies (e.g., organization and critical thinking), metacognitive strategies (e.g., goal setting, monitoring, reflection, and regulation), and resource management strategies (e.g., effort regulation, peer learning, and help seeking). All items were rated on a five-point Likert scale ranging from strongly disagree to strongly agree.

Participants were 41 students enrolled in a general education programming course at a university in northern Taiwan. Survey data were collected at the beginning and end of the semester. A total of 23 valid responses were obtained at both time points.

4.3. Findings

The results showed significant improvements in several aspects of learning motivation after the course. Students' intrinsic goal orientation significantly increased ($t = 3.20$, $p = .0045$), indicating that they were more motivated by personal interest and growth. Self-efficacy also showed a significant increase ($t = 2.61$, $p = .0168$), suggesting greater confidence in their learning ability. Although extrinsic goal orientation, task value, and control of learning beliefs showed slight increases, these changes were not statistically significant. Overall, the Scrum-based instructional approach appeared to enhance students' intrinsic motivation and self-efficacy.

Regarding learning strategies, several dimensions also improved. Among cognitive strategies, elaboration significantly increased ($t = 2.23$, $p = .0117$), indicating that students were more likely to deepen understanding by connecting and explaining concepts. No significant changes were found in overall metacognitive self-regulation. However, within resource management strategies, effort regulation showed a significant improvement ($t = 4.38$, $p = .0006$). This suggests that students became more persistent and resilient when facing challenges, likely supported by the iterative feedback and revision processes embedded in Scrum activities.



(a) Learning Motivation (b) Learning Strategies

Figure 1. The Effects on Self-regulated Learning

5. Concluding Remarks

This study examined the impact of integrating Scrum-based Agile development into a programming course on students’ learning motivation and learning strategies. The results showed improvements in intrinsic goal orientation, self-efficacy, elaboration, and effort regulation, suggesting that iterative feedback and structured teamwork may support both engagement and persistence in learning.

These findings indicate that Agile pedagogy can function not only as a project management approach but also as a learning design that promotes motivational and strategic development. By embedding reflection, accountability, and collaboration into the learning process, Scrum may foster self-regulated learning in programming education. Given the one-group design and limited sample size, future research should adopt comparative or longitudinal approaches to further examine the long-term effects of Agile-based instruction.

Acknowledgements

This project was supported by the Ministry of Education (MOE), Taiwan (MOE-113-TPRGE-1028-003Y1).

References

Fernandes, S., Dinis-Carvalho, J., & Ferreira-Oliveira, A. T. (2021). Improving the performance of student teams in project-based learning with scrum. Education Sciences, 11(8), 444.

Hron, M., & Obwegeser, N. (2022). Why and how is Scrum being adapted in practice: A systematic review. Journal of Systems and Software, 183, 111110.

Klopp, M., Gold-Veerkamp, C., Abke, J., Borgeest, K., Reuter, R., Jahn, S., ... & Landes, D. (2020). Totally different and yet so alike: Three concepts to use scrum in higher education. In Proceedings of the 4th European Conference on Software Engineering Education (pp. 12-21).

Linden, T. (2018). Scrum-based learning environment: Fostering self-regulated learning. Journal of Information Systems Education, 29(2), 65-74.

Linos, P. K., Rybarczyk, R., & Partenheimer, N. (2020). Involving IT professionals in Scrum student teams: An empirical study on the impact of students' learning. In 2020 IEEE Frontiers in Education Conference (FIE) (pp. 1-9). IEEE.

Pintrich, P. R., Smith, D. A., Garcia, T., & McKeachie, W. J. (1993). Reliability and predictive validity of the Motivated Strategies for Learning Questionnaire (MSLQ). Educational andpsychological measurement, 53(3), 801-813.

Vogelzang, J., Admiraal, W. F., & Van Driel, J. H. (2020). A teacher perspective on Scrum methodology in secondary chemistry education. Chemistry Education Research and Practice, 21(1), 237-249.

Hsu, T. C., & Ogata, H. (Eds.). (2026). *Proceedings of the 10th International Conference on Artificial Intelligence, Computational Thinking and STEM Education* .Kumamoto: The Asia-Pacific Society for Computers in Education.

Generative AI Use Patterns in High School Programming and Their Associations with Learner Characteristics

Koki TAMAKI[1*], Akira MAEDA[2] & Shu YAMAMOTO[3]
[1] Research Center for Higher Education, Tokushima University, Japan
[2] Graduate School of Systems Information Science, Future University Hakodate, Japan
[3] Kamiyama Marugoto College of Design, Technology, and Entrepreneurship, Japan
*tamaki.kouki@tokushima-u.ac.jp

Abstract: *Generative AI is increasingly used to support programming learning, yet its educational value may depend on how learners engage with AI and on learner characteristics. This study examined (RQ1) how Japanese high school students used a generative AI coach during micro:bit programming and (RQ2) how pre-survey measures (programming self-efficacy, AI-use frequency, and programming experience) related to observed use patterns. Participants were 122 first-year students in a required informatics course, who could use a browser-based AI system voluntarily across a six-session unit. All interactions were logged and coded for functional purpose and use quality following established schemes (Groothuijsen et al., 2024; Yang et al., 2024). Use was dominated by Solution code generation, with Conceptual understanding as the second most frequent category, while Debugging and Code explanation were rare and Optimization was not observed. Non-ideal/Unproductive use was more frequent overall, but learners varied substantially, and learning-oriented users more often connected generation requests to conceptual checks. Spearman correlations suggested that learner characteristics related more to how students used AI than to whether they used it: programming experience was linked to less reliance on code generation and more conceptual-understanding use, while programming self-efficacy showed a negative relationship with learning-oriented use, underscoring the need for supports that promote conceptual checking and verification across diverse learners.*

Keywords: Generative AI, Programming education, Learning analytics, micro:bit

1. Introduction

Computational Thinking (CT), encompassing decomposition, abstraction, algorithmic design, verification, and debugging, has been emphasized across educational contexts including K–12 (Grover & Pea, 2013), and programming is widely treated as a key means of fostering it. Generative AI has rapidly entered programming education through chatbots, generative models, and intelligent tutoring systems (ITS), offering immediate and personalized feedback (Elnaffar et al., 2026). However, its learning effects are mixed: case studies report concerns about code quality and overreliance (Groothuijsen et al., 2024; Yang et al., 2024), and a quasi-experimental high school study further found lower flow, self-efficacy, and achievement in a ChatGPT-use group (Yang et al., 2025).

AI use likely also varies across learners. Self-regulated learning research in programming documents diverse strategies, the role of metacognition, and learners' difficulties in organizing their thinking, implying that uniform instruction risks over- or under-support (Ramírez-Echeverry et al., 2025). Yet how learners with different characteristics actually use generative AI remains underexamined. To inform interventions that promote educationally desirable use, this study addresses:

RQ1: How do learners use generative AI during programming tasks?

RQ2: What relationships exist between learner characteristics and generative AI use during programming tasks?

2. Methodology

2.1. Participants

A total of 122 first-year students from a Japanese private high school, all enrolled in the required subject Information I, participated. Students were informed that participation would not affect grades and that data would be anonymized and used only for academic purposes. Written informed consent was obtained for both questionnaire and interaction-log data; only consenting students were analyzed.

2.2. Procedure

The study ran during regular Information I classes in October–November 2025 across six sessions using the micro:bit platform. Session 1 provided an orientation (goals, assessment, assignments, environment); Sessions 2–5 were devoted to worksheet-based programming with iterative execution and revision; Session 6 covered completion, submission, and reflection. Generative AI use was not mandated but permitted voluntarily, and was mainly observed in Sessions 2–6 during problem solving and implementation.

2.3. Generative AI Tool and Agent Design

We developed a browser-based conversational system (Google Gemini API, gemini-2.5-flash) that preserves within-session context and stores session-level logs so learners can revisit prior exchanges. To avoid the direct answer-giving typical of generic chatbots, a system prompt configured the agent as a "micro:bit programming coach for high school students," emphasizing scaffolded, step-by-step support that prioritized algorithmic reasoning and built-in functionality over external libraries. The agent was instructed to present solution structures and partial examples with portions left for learners to complete, to provide structural, logical, syntactic, and algorithmic hints as needed, to limit code per message, and to mark unfinished parts with TODO comments, thereby supporting learners' reasoning rather than substituting for it and discouraging copy-and-paste use.

2.4. Analytical Framework

In Session 1 we administered a pre-questionnaire on programming self-efficacy, AI-use frequency, and programming experience, and collected system logs of learners' AI inputs and outputs. Interactions were first coded with Groothuijsen et al.'s (2024) six functional categories: Code explanation (plain-language explanations of existing code), Solution code generation (draft solutions, structures, or pseudocode-to-code from task goals), Error checking and debugging (diagnosing and fixing errors/bugs), Conceptual understanding (deepening concept understanding or comparing implementation options), Solution code optimization (improving existing code), and Mathematical problem solving (computations used in code). Use quality was additionally coded with Yang et al.'s (2024) framework, distinguishing Non-ideal/Unproductive use — shifting cognitive and verification burden to the AI (e.g., minimal self-attempts, delegated diagnosis) — from Learning-oriented use — preserving understanding and autonomy (e.g., substantive attempts, follow-up questions clarifying rationale, rewriting rather than copy-pasting). We then related these codes to the pre-survey measures.

3. Results

3.1. Descriptive Patterns of Generative AI Use

Tables 1 and 2 summarize the labeled interactions (n = 93 learners with ≥ 1 label).
Across functional categories, Solution code generation dominated (370), followed by Conceptual understanding (72); Code explanation and Debugging each appeared about 20 times, Mathematical problem solving once, and Optimization

was not observed. For use quality, Non-ideal/Unproductive use (351) exceeded Learning-oriented use (230; ~39.6%). In both cases, large variances indicate substantial between-learner differences.

Table 1 *Functional categories (Groothuijsen et al., 2024)*

Category	*Total*	*Mean*	*Var*
Code explanation	22	0.24	0.77
Solution code generation	370	3.98	4.18
Error checking and debugging	21	0.23	1.17
Conceptual understanding	72	0.77	1.78
Solution code optimization	0	0.00	0.00
Mathematical problem solving	1	0.01	0.10

n = 93

Table 2 Quality of use (Yang et al., 2024)

Category	*Total*	*Mean*	*Var*
Non-ideal / Unproductive use	351	3.77	3.92
Learning-oriented use	230	2.47	4.07

n = 93

To examine individual differences, we compared learners with Learning-oriented ratios ≥ 0.50 (high group, n = 13) and ≤ 0.20 (low group, n = 7), both with sufficient labels, as exploratory cutoffs. The high group showed frequent co-occurrence of Conceptual understanding with Solution code generation and some Code explanation and Debugging, suggesting back-and-forth interactions involving understanding checks and rationale-seeking rather than stopping at the generated solution. The low group was strongly concentrated in Solution code generation with little Conceptual understanding; learners often pasted code fragments with minimal context, yielding interactions of limited informational content. Thus, effective AI use may depend less on whether learners use AI and more on whether they connect generation to conceptual understanding and articulate their intentions in ways that elicit appropriate support.

3.2. Associations Between Learner Characteristics and Generative AI Use Patterns

Based on the above, correlational analyses focused on three log-derived indices — Solution code generation ratio, Conceptual understanding ratio, and Learning-oriented use ratio — capturing reliance on generation, engagement in conceptual understanding, and overall use quality. Spearman rank correlations with the three pre-survey characteristics are shown in Table 3.

Programming experience was negatively correlated with the Solution code generation ratio ($\rho = -0.26$, $p < .05$) and positively with the Conceptual understanding ratio ($\rho = 0.24$, $p < .05$). Programming self-efficacy was negatively correlated with the Learning-oriented use ratio ($\rho = -0.23$, $p < .05$). AI-use frequency showed only weak, non-significant associations. These results suggest that learner characteristics relate not to whether learners use AI but to how they allocate use across functions. Less experienced learners appear to rely on generation to externalize intended solutions, whereas more experienced learners proceed

Table 3 *Spearman correlations between learner characteristics and generative AI use*

Predictor	Solution code generation ratio	Conceptual understanding ratio	Learning-oriented use ratio
Programming self-efficacy	0.03	-0.03	-0.23*
AI use frequency	-0.12	0.21	0.09
Programming experience	-0.26*	0.24*	0.14

*$p < .05$

more independently and use AI mainly to clarify uncertainties and verify rationales. AI-use frequency alone does not guarantee desirable use. The negative association between self-efficacy and Learning-oriented use further suggests divergent help-seeking: high-efficacy learners may use AI instrumentally via concise prompts for efficiency, while lower-efficacy learners may rely on AI for conceptual scaffolding through more exploratory dialogues.

4. Conclusion

This study described high school students' generative AI use during micro:bit programming and its associations with learner characteristics. For RQ1, use was dominated by Solution code generation, followed by Conceptual understanding, with Debugging and Code explanation rare and Optimization absent; Non-ideal/Unproductive use was more frequent overall, but with substantial individual variation, as higher Learning-oriented learners more often connected generation to conceptual checks. For RQ2, programming experience was negatively associated with the Solution code generation ratio and positively with the Conceptual understanding ratio; AI-use frequency showed only weak associations; and programming self-efficacy was negatively associated with the Learning-oriented ratio, warranting further examination. Overall, AI use varied with learner characteristics, suggesting that appropriate use does not emerge uniformly. Limitations include the single-school, short-term setting, exploratory cutoffs, and correlational design. Practically, the results imply a need for differentiated guidance: encouraging conceptual verification among high-efficacy learners and supporting novices in articulating goals before requesting code. Future work should design adaptive supports that reduce overreliance and promote conceptual checking and verification across diverse learner profiles.

References

Elnaffar, S., Rashidi, F., & Abualkishik, A. Z. (2026). Teaching with AI: A systematic review of chatbots, generative tools, and tutoring systems in programming education. International Journal of Learning, Teaching and Educational Research, 25(1), 1–28.

Groothuijsen, S. E. A., van den Beemt, A., Remmers, J. C., & van Meeuwen, L. W. (2024). AI chatbots in programming education: Students' use in a scientific computing course and consequences for learning. Computers and Education: Artificial Intelligence, 7, 100290.

Grover, S., & Pea, R. (2013). Computational thinking in K–12: A review of the state of the field. Educational Researcher, 42(1), 38–43.

Ramírez-Echeverry, J. J., Restrepo-Calle, F., & Torres Jiménez, S. (2025). Self-regulated learning strategies in computer programming education. European Journal of Education, 60, e70052.

Yang, S., Zhao, H., Xu, Y., Brennan, K., & Schneider, B. (2024). Debugging with an AI tutor: Investigating novice help-seeking behaviors and perceived learning. In ACM Conference on International Computing Education Research V.1 (ICER '24 Vol. 1) (11 pages). Association for Computing Machinery.

Yang, T.-C., Hsu, Y.-C., & Wu, J.-Y. (2025). The effectiveness of ChatGPT in assisting high school students in programming learning: Evidence from a quasi-experimental research. Interactive Learning Environments, 33(6), 3726–3743.

Hsu, T. C., & Ogata, H. (Eds.). (2026). *Proceedings of the 10th International Conference on Artificial Intelligence, Computational Thinking and STEM Education* .Kumamoto: The Asia-Pacific Society for Computers in Education.

STEM Education: Benefits and Challenges of Using the Demonstration Method

Chak-Him FUNG[1], Siu-Ping NG[2*]
[1] Hong Kong Baptist University, Hong Kong SAR, China
[2] The Education University of Hong Kong, Hong Kong SAR, China
*s0655050@alumni.eduhk.hk

Abstract: *The education field aims to develop STEM teaching methods that integrate knowledge across disciplines and connect it to daily life. While practical work helps students grasp abstract concepts, teachers often avoid it due to resource constraints and students' lack of experimental skills, opting for demonstrations instead. This study systematically reviewed 49 studies, identifying 15 benefits and 14 challenges of utilizing the demonstration method in STEM education. Recommendations are provided for policymakers, universities, and schools to improve demonstration-based teaching outcomes.*

Keywords: Demonstration, effectiveness, efficacy, STEM education

1. Introduction and Literature Review

STEM education, encompassing science, technology, engineering, and mathematics, has gained prominence due to its strategic importance and labor market demand (Akcan et al., 2023). However, educators face significant challenges in implementing effective STEM teaching, such as connecting theoretical knowledge to real-world applications (Hong Kong Federation of Education Workers, 2017). Practical work is often recommended to address these issues (Abrahams & Reiss, 2010; Fung, 2020), as it engages students in active learning and skill development. Yet, barriers like insufficient resources and students' limited prior knowledge hinder its implementation. As an alternative, the demonstration method (DM) is frequently employed (Moore et al., 2020), using tangible tools or experiments to illustrate concepts and procedures. The DM is supported by theories like Cognitive Load Theory (Sweller et al., 2019) and Constructivism (Hmelo-Silver et al., 2007), which highlight its potential to enhance understanding, engagement, and academic performance. However, its effectiveness remains debated, with studies showing mixed results depending on the context and implementation.

Given the widespread use of the DM in STEM education and its potential to address pedagogical challenges, it is crucial to examine its effectiveness. While the DM offers benefits such as reducing cognitive load and fostering collaborative learning, inconsistent findings in existing research highlight the need for a systematic review. This review aims to explore the mechanisms of the DM, particularly its benefits and challenges, and provide insights to improve STEM teaching practices.

2. Research Questions

To achieve these objectives, this study seeks to address the following research questions: What are the benefits and challenges of utilizing the demonstration method (DM) in STEM education?

3. Methodology

The search process utilized the Scopus database with keywords related to the DM approach, such as "demonstration method," "teacher demonstration," and "teaching demonstration," yielding 853 results as of 2 December 2024. Titles, abstracts, and keywords were scanned, and non-English items and non-article publications were excluded, leaving 482 articles. Abstracts were then reviewed, and materials unavailable in full-text format online were removed. Articles unrelated to the DM but containing key terms in their abstracts were also excluded, resulting in 180 articles after eliminating duplicates. Further filtering excluded non-empirical studies, non-STEM-related research, and one study due to an inaccessible website flagged for security risks. Ultimately, 49 publications met the criteria and were included in this review.

4. Results

The demonstration method offers 15 benefits, including promoting knowledge and understanding, enhancing academic performance, and fostering interest, motivation, and positive learning attitudes. Students also show a preference for this method, with additional advantages such as cross-disciplinary linkage, development of technical lab skills, reduced cognitive load, increased curiosity, and mitigating the impact of inadequate lab skills. For faculty, it provides comparable effects to other strategies, is favorable to teachers, increases student attention, and addresses blind spots. Operationally, it saves time and improves preparation efficiency.

The demonstration method faces 14 challenges, including limited impact on academic performance, procedural knowledge, lab skills, soft skills, and science literacy, as well as reduced opportunities for mistakes and a lack of understanding of purpose and connections. Faculty-related issues include teaching difficulties, challenges in checking learning progress, lack of professional training, and limited immediate feedback. Operationally, constraints such as limited teaching hours, insufficient materials or equipment, and difficulties in managing large class sizes also pose significant challenges.

5. Conclusion and Implications

This systematic literature review provides an in-depth analysis of the DM in STEM education. This review analyzes the use of DM in STEM education, summarizing 49 studies that highlight 15 benefits and 14 challenges affecting students, teachers, and operational aspects. By understanding these insights, we aim to support both educators and students, fostering more effective teaching practices in STEM education.

This study provides recommendations for policymakers, universities, and schools to enhance the use of the DMs. Policymakers should allocate resources for professional development programs on DMs, fund additional staff to assist with demonstration preparation, and support curriculum planning. Universities can expand teacher training by offering courses to equip educators with skills to implement DMs effectively. Schools should organize collaborative sessions for teachers to share experiences and foster a supportive environment, empowering them to improve DM implementation.

References

Abrahams, I., & Reiss, M. (2010). Effective practical work in primary science: The role of empathy. Primary Science, 113, 26–27. https://discovery.ucl.ac.uk/id/eprint/10024060/1/Reiss_Abrahams%20Reiss%202010%20Primary%20Science.pdf

Akcan, A. T., Yıldırım, B., Karataş, A. R., & Yılmaz, M. (2023). Teachers' views on the effect of STEM education on the labor market. Frontiers in Psychology, 14, 1184730. https://doi.org/10.3389/fpsyg.2023.1184730

Fung, C. H. (2020). How does flipping classroom foster the STEM education: A case study of the FPD model. Technology, Knowledge and Learning, 25(3), 479–507. https://doi.org/10.1007/s10758-020-09443-9

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). Educational Psychologist, 42(2), 99–107. https://doi.org/10.1080/00461520701263368

Hong Kong Federation of Education Workers. (2017). Research report of support policy to the frontline STEM teacher. https://hkfew.org.hk/en/%E6%95%99%E8%82%B2%E6%94%BF%E7%AD%96/item/download/1868_decc5f308549ea804db32ab7dbccbfa8

Moore, A. M., Fairhurst, P., Correia, C., Harrison, C., & Bennett, J. (2020). Science practical work in a COVID-19 world: Are teacher demonstrations, videos and textbooks effective replacements for hands-on practical activities? School Science Review, 102(378), 7–12.

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

Hsu, T. C., & Ogata, H. (Eds.). (2026). *Proceedings of the 10th International Conference on Artificial Intelligence, Computational Thinking and STEM Education* .Kumamoto: The Asia-Pacific Society for Computers in Education.

AI-Powered MCQ Generation: Enhancing Teaching Efficiency and Student Learning

Xiajuan YE [1*], Sheau Huey LEE[2], Yu Ming CHAN[3] & Teck Choow LIM[4]
[1,2,3,4] NUS High School of Math and Science, Singapore
*nhsyx@nus.edu.sg

Abstract: *Unlike what they had in lower years, there are no practice worksheets during lessons for the Year 5 students taking Advanced Calculus before they sit for graded tasks. Teachers sought to use AI-generated questions to better prepare students while addressing the fundamental challenge of knowing whether students have learned and how to respond when they need additional guidance. This study explores AI-powered MCQ generation as a strategy to help students gain mastery without significantly increasing teacher workload. We implemented two approaches: Student Learning Space Authoring Copilot for generating MCQs with auto-marking and feedback features across two curriculum chapters, and Microsoft Copilot/ChatGPT with Microsoft Forms' auto-marking capabilities for another two chapters. Although teachers are still wary of questions generated by AI tools, with professional expertise in content knowledge and appropriate intervention, we may still use them to lighten the burden on teachers and focus on other lesson parts. Survey results from students show that 72% of the students think positively regarding AI in helping them in their learning, with some commented that AI tool turned out to be more useful that what they used to think.*

Keywords: Artificial Intelligence (AI), ChatGPT, Microsoft Copilot, Student Learning Space (SLS), Authoring Copilot (ACP), Microsoft Forms (MS Forms)

1. Introduction

The big idea behind this is how we know if students have learned and how we respond when they need help from teachers. We wish to use a variety of data to assess the impact of methods used on students' learning and explore other strategies beyond the usual mode of learning to help students gain mastery without adding too much extra workload to teachers. Teachers notice that there have been no practice worksheets to prepare students for the graded tasks. We hope that AI-generated questions can be used to better prepare students for graded assessments. We aim to see how artificial intelligence in education can be used to support and meet the needs of teaching and learning.

As teachers may have their own preference, we allowed different explorations to happen within the same teaching and learning framework. Two teachers used Authoring Copilot in Student Learning Space (SLS) for generating MCQs with auto-marking and feedback feature for two chapters while the other two teachers used Microsoft Copilot/ChatGPT with Microsoft Forms'(MS Forms) auto-marking capabilities for another two chapters. These generated questions were used as formative assessments to support teaching and learning. The four chapters are Limits, Differentiation, Integration and Differential Equations from Chapter 1 to Chapter 4 respectively.

As a specialised independent school, the students in our school are considered high ability learners in the country. The MCQs generated were assigned to 175 high ability learners from the Year 5 cohort. The questions were used as formative quizzes and open for students to attempt one week before each graded task.

2. Auto-marking and/or feedback in SLS Authoring Copilot and MS Forms

2.1. Auto marking and customised feedback in SLS Authoring Copilot

Authoring Copilot (ACP) in Singapore's SLS is an AI-powered tool that assists teachers in quickly creating and customizing digital lesson content, like quizzes, sections, and activities, by turning simple text prompts into structured learning materials, significantly streamlining lesson planning. In our practice, Authoring Copilot in SLS was used to create quizzes. MCQs were generated for each quiz. For MCQs, questions generated allows auto-marking, which saves significant time for teachers and gives instant feedback to students. What benefits students more is that the MCQs generated in SLS also allow response-specific feedback. This gives not just hint but also helps students close some learning gaps they face in their studies. The screenshots below give an example of what response-specific feedback looks like.

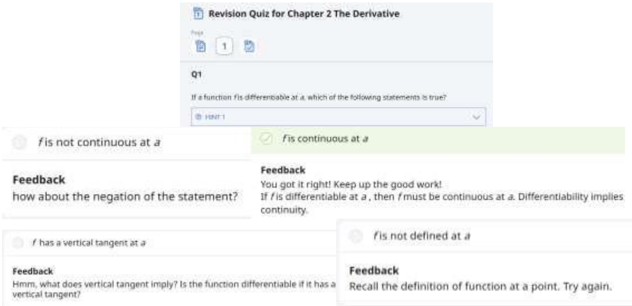


Figure 1. The auto-marking and response-specific feedback for MCQs in SLS

2.2. Auto-marking in MS Forms

MS Forms is a simple web-based application for creating and sharing online surveys, quizzes, polls, and forms to collect feedback or test knowledge, accessible on any device, with features for real-time results, built-in analytics, and easy export to Excel for deeper analysis. Multiple-choice questions were generated using ChatGPT, and MS Forms served as the hosting platform where the questions were deployed and distributed to students via shareable links.

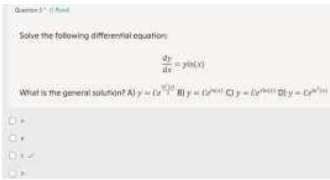


Figure 2. Auto-marking of MCQs in MS Forms

3. Result analysis

3.1. Quantitative result analysis

A couple of days before the actual assessment, teachers can access different platforms to check students' submission. This will help to identify students who need help but may not step forward to ask for help. Teachers then can check with the students to arrange necessary remedials before they sit for the graded tasks.

In this section, we analysed the result of each quiz to see how well the students did the quizzes. When entering the prompts, we specified our requirement such as 30% simple questions, 50% intermediate level questions and 20% challenging questions.

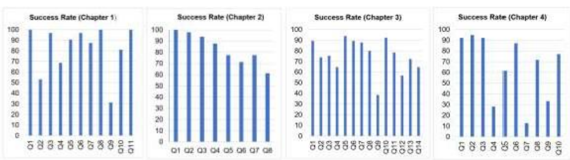


Figure 3. Quantitative Results for All Chapters

From the results above, teachers can quickly identify questions that students are weak in general in respective chapters. With this, teachers can address the relevant learning points to the students. Teachers can also dive into individual students' results to cater to their learning needs.

On AI generated questions, for chapters 1, 3 and 4, we can see the success rate spread ranges from around 64% to 72%. This indicates that the effectiveness of AI-generated could be dependent on specific question characteristics or content areas. For chapter 2, the difficulty level of the questions increases gradually. And questions generated for chapter 4 may also suggest that although half the questions were done well by students, the unpredictable nature of performance makes AI-generated questions less reliable as a standalone tool. Teachers cannot assume consistent quality across different question types. We can see that AI as a valuable supplementary tool that can significantly enhance teacher productivity in question creation whilst maintaining the need for professional judgement in final selection and refinement.

3.2. Qualitative result analysis

A perception survey was conducted after all the formative quizzes were done. There were five multiple choice questions and three free response questions. The following questions were asked.

Q1: The AI-tools have successfully helped me achieve the lesson's learning outcomes. Q2: I am capable of working on all AI-based activities.

Q3: Compared to traditional lessons, the AI-tools have enhanced my learning experience.

Q4: Compared to traditional lessons, the AI-tools have motivated me to take a more active role in the learning process.

Q5: Before using the AI-tools during the lesson(s), what was your perception of the AI tool?

Q6: How has the adoption of the AI-tools in the lesson(s) changed your perspective, and what specific benefits have you experienced in your learning process as a result?

Q7: Were there any difficulties or challenges you faced while using the AI-tool? Describe briefly.

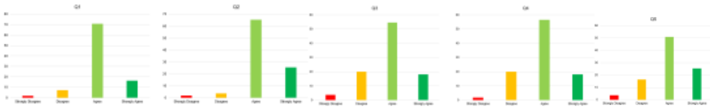


Figure 4. Qualitative Results for the First Five Questions

On average, 80.28% of the students chose "agree" or "strongly agree" for the first five questions. This shows that the AI tools have helped and motivated students to play a more active role in their learning. As for teachers, the AI-tools were efficient in creating questions up to a certain level which helped reduce teachers time in preparation.

When asked what their perception was before using the AI-tools, students answered that AI-tools are perfect and give reliable answers and the tools can be used to enhance their learning or provide them with practice questions.

When asked how the adoption of the AI tools changed their perspective and the benefits they experienced, students responded that AI tools are more useful than what they originally thought, and they can use the tools to enhance their learning. Some students also commented that it would be easier for teachers to make new questions to test their knowledge. When asked about the difficulties or challenges they faced while using the AI-tools, students realized that AI sometimes gave incorrect answers. A few students pointed out that the hallucinations set them back in understanding the questions and how to solve them.

4. Discussion

In the perception survey, more than 72% of the students think positively regarding AI in helping them in their learning, with some commented that AI tools turned out to be more useful that what they used to think.

Some MCQs have more than one acceptable answer, but the system only accepts one of them as “correct”. Some students will take the initiative to clarify with teachers if they think that the questions/answers are inaccurate. AI tools have (unintentionally) motivated them in their critical thinking even though there are mistakes in the questions generated by AI tools. Besides errors in questions, questions generated are quite basic and much easier than expected given that our students are talented. Thus, teachers are to modify the questions to address the learning needs of students.

Although teachers are still wary of questions generated by AI tools, with professional expertise in content knowledge and appropriate intervention, we may still use them to lighten the burden on teachers and focus on other lesson parts. The analysis allowed teachers to take a deeper look at the students’ responses and use it as a guideline for future lesson preparation and pedagogical purposes or to help students before graded assessments.

Acknowledgements

We would like to thank our school for providing us with the platform and the opportunity to explore innovations in teaching.

References

Yavuz S.K. & Emre Emekli. (2024). ChatGPT prompts for generating multiple-choice questions in medical education and evidence on their validity: a literature review. Postgraduate Medical Journal, 100 (1189), 858–865.

Muhammad Y.M., Ahmed Tlili, Georgios Lampropoulos & Ronghuai Huang. (2024). A systematic review of literature reviews on artificial intelligence in education (AIED): a roadmap to a future research agenda. Smart Learning Environments, 11(1), 1-33.

Hsu, T. C., & Ogata, H. (Eds.). (2026). *Proceedings of the 10th International Conference on Artificial Intelligence, Computational Thinking and STEM Education* .Kumamoto: The Asia-Pacific Society for Computers in Education.

The impact of Integrating Python and AI into Discipline-Related Courses on the Computational Thinking Skills of Students in Mold and Die Engineering

Mingyu LIAO[1*], Chisheng ZHOU [2], Yurou ZHENG [3], Kaiting ZHAN [4], KaiSheng ZHAN [5] Wan-Ju LIN [6], Yu-Ping HSU [7] & Kuang-Chen HSU [8]

[1,2,3,4,5] Department of International Business, National Kaohsiung University of Science and Technology, Taiwan

[6] Department of Mold and Die Engineering, National Kaohsiung University of Science and Technology, Taiwan

[7] School of Medicine, University of Kansas Medical Center, USA

[8] Learning's Office of Digital Learning, University of Notre Dame, USA

*mygliao@nkust.edu.tw

Abstract: *In the era of Industry 4.0 and AI, the manufacturing sector is undergoing a digital transformation. Traditional mold and die engineering now requires talent proficient in cross-disciplinary integration and complex problem-solving rather than just machining skills. Consequently, fostering digital literacy and critical thinking has become a priority in educational reform.*

For mold engineering students, computational thinking is now a core competency. Integrating Python and AI allows students to automate data processing while enhancing structured thinking through algorithmic training. However, effectively embedding these tools into traditional curricula and measuring their impact remain challenges.

This study examines the impact of AI-integrated learning in mold and die engineering across four dimensions: creativity, algorithmic thinking, collaboration, and critical thinking. Through empirical analysis, the research evaluates technological interventions and the role of instructional design, aiming to provide pedagogical recommendations for the digital evolution of engineering education.

Keywords: AI-integrated learning, Computational Thinking (CT), Python, Algorithmic Thinking, Engineering Education

1. Introduction

The era of Industry 4.0 and Artificial Intelligence (AI) has pushed the manufacturing sector toward digital transformation. In mold and die engineering, traditional machining skills must now be supplemented with cross-disciplinary integration and complex problem-solving.

Consequently, Computational Thinking (CT)—encompassing digital literacy and algorithmic logic—has emerged as a core competency. This study explores the impact of an 18-week Python and AI-integrated curriculum on undergraduate students, focusing on four dimensions: creativity, algorithmic thinking, collaboration, and critical thinking.

2. Literature Review

2.1. Computational Thinking (CT) in AI Education

CT involves problem decomposition, abstraction, and algorithmic design (Wing, 2006). Integrating Python programming and AI literacy is essential for modern problem-solving (Kong et al., 2021). Research suggests that AI-integrated programming courses may enhance learning outcomes and cognitive development (Sung et al., 2017).

2.2. Self-Assessment Framework

This study utilizes the Computational Thinking Scale (CTS) developed by Korkmaz et al. (2017). The CTS provides a validated, multi-dimensional assessment of students' self-perceived computational thinking abilities across Creativity, Algorithmic Thinking, Cooperativity, and Critical Thinking.

3. Research method

3.1. Research Design and Participants

This study employed a quasi-experimental one-group pretest-posttest design to evaluate the learning effectiveness of the proposed course. The participants were 94 undergraduate engineering students in Taiwan. All participants were enrolled in a programming course integrating Python and Artificial Intelligence (AI). A pre-test was delivered in the first week of the 18-week course, and a post-test using the same instrument was conducted at the end of the semester to evaluate the changes in their computational thinking.

3.2. Instrument

The questionnaire consisted of 29 items comprising four dimensions: Creativity (Items 1-8), Algorithmic Thinking (Items 9-14), Cooperativity (Items 15-18), and Critical Thinking (Items 19-29). All items were scored on a 5-point Likert scale, ranging from 1 (strongly disagree) to 5 (strongly agree). Higher scores indicate greater computational thinking ability.

3.3. Data Analysis

The collected data were analyzed using SPSS, and paired-samples t-tests were performed to examine pre- and post-test differences and evaluate the impact of the AI-integrated programming course on students' computational thinking.

4. Results

This study employed a paired-sample t-test to examine the differences between pretest and posttest scores following the implementation of AI- and Python-integrated learning. The overall results indicated significant differences across all four dimensions (p<.05). Effect sizes ranged from small to moderate (See Table 1).

Table 1. *Paired-samples t-test of Pre-test and Post-test for the Four Concepts*

Four Dimensions of Computational Thinking	Mean M (Standard Deviation SD)		*df*	t-value	Two tailed p	Cohen's d
	Pre-test	Post-test				
1. Creativity	3.90 (.607)	3.76 (.564)	93	-2.491	.015*	-.257
2.Algorithmic Thinking	3.70 (.691)	3.40 (.733)	93	-3.917	<.001***	-.404
3. Cooperativity	3.92 (.701)	3.68 (.739)	93	-3.205	.002**	-.331
4. Critical Thinking	3.55 (.617)	3.36 (.687)	93	-2.273	.025*	-.234

**p<.05 , **p<.01 , ***p<.001*

In general, the posttest mean scores showed a significant decrease compared to the pretest mean scores. Algorithmic thinking showed the largest decline.

5. Discussion

The findings revealed that the posttest mean scores across all four dimensions were significantly lower than the pretest scores. The instructional interventions are usually expected to improve ability levels; however, decreases in self-assessment are not uncommon in educational research. Drawing on the Dunning–Kruger effect (1999), these results may reflect students' development of a more cautious, realistic self-assessment after exposure to AI and Python tools. Prior to the course intervention, students may have held relatively high subjective evaluations of their creativity and thinking abilities based on prior experiences or intuitive judgments. However, after engaging in problem-solving, comparing alternative solutions, and conducting systematic analyses using AI and Python tools, students may have gained a clearer understanding of the complexity and limitations of their abilities, leading them to adjust their self-evaluation standards and resulting in lower posttest scores.

This study showed that the effect sizes were small to moderate. Although all differences were statistically significant, the magnitude of change was modest. This may be related to factors such as the duration of the intervention, students' familiarity with AI and Python, or the instructional design and guidance strategies employed in the course.

Overall, the results suggest that AI- and Python-integrated learning may promote a deeper reconsideration and reconstruction of their thinking processes rather than increasing self-assessment scores.

Acknowledgements

We would like to thank everyone who has supported this article.

References

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. Journal of Personality and Social Psychology, 77(6), 1121–1134. https://www.researchgate.net/publication/12688660_Unskilled_and_Unaware_of_It_How_Difficulties_in_Recognizing_One's_Own_Incompetence_Lead_to_Inflated_Self-Assessments

Korkmaz, Ö., Çakır, R., & Özden, M. Y. (2017). A validity and reliability study of the computational thinking scales (CTS). Computers in Human Behavior, 72, 558-569.

Kong, S. C., Abelson, H., Salvatore, C., & Lai, M. (2021). Artificial intelligence literacy: Definition, teaching, and learning. Computers and Education: Artificial Intelligence, 2, 100021. https://doi.org/10.1016/j.caeai.2021.100021

Lye, S.Y., & Koh, J.H.L. (2014). Review on teaching and learning of computational thinking through programming: What is next for K-12? Computers in Human Behavior, 41, 51-61.

Sung, H. Y., Chang, J.H., & Lee, M.D. (2017). The effects of integrating mobile devices with teaching and learning on students' learning performance: A meta-analysis and research synthesis. Computers & Education, 94, 252-275.

Wing, J.M. (2006). Computational thinking. Communications of the ACM, 49(3), 33-35.

Younis, B. (2025). The artificial intelligence literacy (AIL) scale for teachers: A tool for enhancing AI education. Journal of Digital Learning in Teacher Education, 41(1), 37-56.

Hsu, T. C., & Ogata, H. (Eds.). (2026). *Proceedings of the 10th International Conference on Artificial Intelligence, Computational Thinking and STEM Education* .Kumamoto: The Asia-Pacific Society for Computers in Education.

Judgment Delegation and Responsibility Attribution in Autonomous Driving Robot Learning

Sojung AHN[1], Jeongmin LEE[2*]
[1,2] Department of Educational Technology, Ewha Womans University, Korea
ahnsojung@ewha.ac.kr, *jeongmin@ewha.ac.kr

Abstract: *This study examines how students' judgment delegation and responsibility attribution change in a human–AI collaborative learning environment. To address the limitations of AI ethics education focused primarily on attitudes, a three-session autonomous driving robot program using LEGO Education SPIKE Prime was implemented with 17 upper elementary students. Data were collected through pre- and post-surveys and activity sheets. The results show that students shifted from uncritical acceptance of AI decisions toward reviewing and verifying AI outputs. Responsibility attribution increased most notably for users, while responsibility attributed to AI systems and developers also increased and responsibility attributed to manufacturers slightly decreased. These changes were associated with experiences of AI errors and reflective activities during learning. The findings suggest that AI ethics education should emphasize experience-based learning design, supporting students' critical engagement with AI and their awareness of responsibility in human–AI collaboration.*

Keywords: Autonomous driving robot learning, Human–AI collaboration, Judgment delegation, Responsibility attribution, Learning experience design

1. Introduction

The rapid advancement of generative AI and automation is transforming education into human–AI collaborative environments, where learners not only use AI but also review and adjust AI's decisions (Kong et al., 2025). The delegation of judgment to AI has been widely reported (Westphal et al., 2023; Albers et al., 2026), yet less is known about how learners attribute responsibility in AI-assisted decision-making contexts.

Recent frameworks (UNESCO, 2023; OECD, 2026) emphasize that AI ethics education should move beyond normative instruction toward experience-based learning, where learners actively make decisions and reflect on responsibility. However, existing studies have largely focused on attitudes, providing limited insight into how judgment delegation and responsibility attribution evolve.

To address this gap, this study aims to examine how students' judgment delegation and responsibility attribution evolve through experience-based learning in a human–AI collaborative environment. A three-session autonomous driving robot program using LEGO Education SPIKE Prime was implemented, and data were collected through pretest and posttest surveys and activity sheets.

2. Method

This study included 17 upper elementary students who participated in a three-session autonomous driving robot learning program using LEGO Education SPIKE Prime. Data were collected through pretest and posttest surveys measuring delegation, oversight, AI trust, and responsibility attribution, as well as activity sheets documenting students' decisions and reasoning.

Descriptive statistics were used to analyze changes in survey responses. Scenario-based responses were coded to identify patterns of judgment delegation and responsibility attribution. Learner patterns were categorized into three types—stable, human-responsibility-oriented, and context-dependent—based on changes across sessions.

3. Result

3.1. Survey Results: Changes in Judgment Delegation

The pretest–posttest survey results indicate a shift in students' judgment delegation. As shown in Figure 1, delegation decreased from pretest (M = 2.20, SD = 0.63) to posttest (M = 1.57, SD = 0.58), suggesting a reduced tendency to accept AI decisions without evaluation. In contrast, oversight increased from pretest (M = 3.00, SD = 0.45) to posttest (M = 3.20, SD = 0.48), indicating a greater tendency to review and verify AI outputs. Meanwhile, AI trust remained relatively stable (pretest: M = 3.74; posttest: M = 3.71), indicating that students maintained their trust in AI while engaging with it more critically.

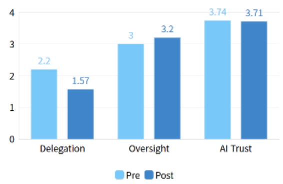


Figure 1. Pre–Post Comparison of Mean Scores in Delegation, Oversight, and AI Trust

3.2. Scenario-Based Results: Changes in Responsibility Attribution

Analysis of scenario-based responses revealed changes in responsibility attribution. As presented in Table 1, responsibility attributed to users increased in both scenarios, while responsibility assigned to AI systems and developers also increased; in contrast, responsibility attributed to manufacturers slightly decreased.

The increase in user responsibility was especially notable, indicating a growing recognition of human accountability in AI-assisted contexts. At the same time, the results do not suggest a simple shift away from AI systems, as responsibility assigned to AI systems also increased in both scenarios. Rather, students appeared to develop a more differentiated understanding of responsibility, placing greater emphasis on proximal human actors, particularly users.

Table 1. *Pre–Post Changes in Mean Responsibility Attribution by Scenario*

	Agent	Pre (M)	Post(M)	Change
S1	User	2.57	3.14	+0.57
	AI System	3.29	3.86	+0.57
	Developer	3.86	4.14	+0.28
	Manufacturer	3.93	3.79	-0.14
S2	User	2.50	3.29	+0.79
	AI System	3.79	4.29	+0.50
	Developer	3.71	3.86	+0.15
	Manufacturer	3.64	3.36	-0.28

3.3. Learner Pattern Analysis

To further examine individual differences, learners were categorized into three patterns: stable, human-responsibility-oriented, and context-dependent. Most students exhibited a human-responsibility-oriented pattern, showing increased emphasis on human accountability. Some maintained consistent perceptions (stable), while others demonstrated context-dependent variation. The human-responsibility-oriented pattern indicates a shift toward actively recognizing human accountability in AI-assisted decision-making. In contrast, the stable pattern reflects consistent perceptions across pretest and posttest, while the context-dependent pattern suggests that learners adjusted their responsibility attribution depending on the situation. These findings suggest that although overall trends were observed, individual patterns of change varied across learners.

Table 2. *Distribution of Responsibility Attribution Change Patterns*

Behavior Change Type	Frequency (n)	Percentage (%)
Stable Responsibility Pattern(CH1)	4	23.5%
Human-Responsibility-Oriented Pattern (CH2)	9	52.9%
Context-Dependent Responsibility Pattern (CH3)	4	23.5%
Total	17	100%

4. Conclusion and Discussion

This study examined changes in students' judgment delegation and responsibility attribution in an autonomous driving robot-based learning environment. The results showed that delegation to AI decreased, while oversight increased, indicating a shift toward more critical engagement with AI. Responsibility attribution changed most notably toward greater user responsibility, while responsibility attributed to AI systems and developers also increased and responsibility attributed to manufacturers slightly decreased. These findings suggest that students developed a more differentiated understanding of responsibility in AI-assisted contexts. Most learners demonstrated a human-responsibility-oriented pattern, with some stable or context-dependent variations. These findings highlight the importance of experience-based AI ethics education, particularly learning designs that include intentional error situations to promote awareness of AI limitations and the need for human intervention.

However, this study is limited by a small sample size and reliance on survey-based data. Future research should incorporate qualitative methods, such as interviews and observations, to provide more in-depth insights.

References

Albers, F., Strauß, S., Rummel, N., & Köbis, N. (2026). Are they just delegating? Cross-Sample Predictions on University Students'& Teachers' Use of AI. arXiv preprint arXiv:2601.21490. https://doi.org/10.48550/arXiv.2601.21490

Kong, X., Fang, H., Chen, W., Xiao, J., & Zhang, M. (2025). Examining human–AI collaboration in hybrid intelligence learning environments: insight from the Synergy Degree Model. Humanities and Social Sciences Communications, 12(1), 1-14. https://doi.org/10.1057/s41599-025-05097-z

Organisation for Economic Co-operation and Development (2026). OECD Digital Education Outlook 2026: Exploring effective uses of generative AI in education. OECD Publishing. https://doi.org/10.1787/062a7394-en

UNESCO. (2023). Guidance for generative AI in education and research. UNESCO Publishing. https://doi.org/10.54675/EWZM9535

Westphal, M., Vössing, M., Satzger, G., Yom-Tov, G. B., & Rafaeli, A. (2023). Decision control and explanations in human-AI collaboration: Improving user perceptions and compliance. Computers in Human Behavior, 144, 107714. https://doi.org/10.1016/j.chb.2023.107714

Hsu, T. C., & Ogata, H. (Eds.). (2026). *Proceedings of the 10th International Conference on Artificial Intelligence, Computational Thinking and STEM Education* .Kumamoto: The Asia-Pacific Society for Computers in Education.

LLM-Driven Feedback in Chemical Risk Management Training Based on Story-Centered Curriculum

Kengo KATAYAMA[1*], Yoshihiro YAMAGUCHI[2], Toshihiro KITA[3], Hiroshi NAKANO[4] & Yoshiko GODA[5]

[1,2] Graduate School of Science and Technology, Kumamoto University, Japan

[3,4,5] Graduate School of Instructional Systems, Kumamoto University, Japan

*257d9363@st.kumamoto-u.ac.jp

Abstract: *Safe handling of chemicals in university laboratories requires students to assess risks and implement appropriate mitigation measures. To support this, the authors previously developed an e-learning course integrating the ARCS-V motivational model and the Story-Centered Curriculum. However, formative evaluation revealed that delays in instructor feedback reduced learner motivation, particularly the "Volition" component. This preliminary study examines the feasibility of using a Large Language Model to provide rapid, personalized feedback on student assignments. A prompt was designed to emulate a laboratory supervisor and incorporate ARCS-V elements—especially Satisfaction and Volition. Fourteen feedback comments were generated and reviewed by a chemical safety expert, who judged them to be accurate, practical, and motivational. Notably, the expert observed that AI-generated feedback sometimes stimulated a stronger sense of accomplishment than human feedback. This study also discusses limitations related to incorrect or unsafe responses and outlines initial mitigation strategies.*

Keywords: LLM, Chemical Risk Management, Story-Centered Curriculum, ARCS-V, Feedback

1. Introduction

University laboratories handle diverse chemicals, requiring students to assess risks and implement appropriate mitigation measures. Although students handling chemical substances typically possess foundational knowledge, they often lack sufficient practical experience for informed safety decision-making, highlighting the need for applied safety competencies. To address this need, the authors developed an e-learning course based on the ARCS-V model (Keller, 2008) and the Story-Centered Curriculum (SCC) (Schank, 2007), which supports learner motivation through Attention (A), Relevance (R), Confidence (C), Satisfaction (S), and Volition (V), and situates learning in realistic task contexts. In this paper, assignments refer to major learner submissions within the course, while tasks denote the individual activities performed within each assignment. However, a previous pilot study revealed a critical bottleneck: delays in instructor feedback for open-ended tasks reduced learner motivation, particularly affecting Volition—the will to persist until course completion. This preliminary study examines the feasibility of using a Large Language Model (LLM) to provide immediate, pedagogically sound feedback and evaluates whether ARCS-V–based LLM-generated feedback is perceived by experts as accurate and motivational.

2. Related Work

Effective feedback must be timely and specific to reduce performance gaps (Hattie & Timperley, 2007). Although LLMs can generate rapid, personalized feedback aligned with instructor criteria (Matelsky et al., 2023), prior studies have focused mainly on structural quality rather than motivational impact (Mazzullo & Bulut, 2024). This study addresses this gap by integrating ARCS-V motivational principles into LLM-generated feedback within an SCC-based chemical risk management course and evaluating its effectiveness through expert review.

3. Methods

Figure 1 shows the SCC learning flow with LLM-driven feedback. Building on this flow, we created a prompt that instructed the LLM to assume the role of a university laboratory supervisor when generating feedback for student assignments. The LLM was instructed to provide feedback as both a subject-matter expert and an educator. For each assignment, task descriptions and evaluation criteria were inserted into the prompt. The LLM was instructed to judge whether each submission met the assignment-specific evaluation criteria. The prompt also required the LLM to follow a fixed feedback structure, including an overall comment, a pass/resubmission judgment with reasons, strengths, improvement advice, and expectations for the next assignment. Feedback was limited to approximately 10 lines for conciseness. Table 1 summarizes the LLM prompt components and their ARCS-V alignment.

The LLM used in this study was OpenAI's GPT-5.2 Instant, accessed through the web interface (https://chatgpt.com). Using the prompt described above, feedback was generated for assignments 1 through 7 submitted by two learners, resulting in a total of fourteen feedback comments.

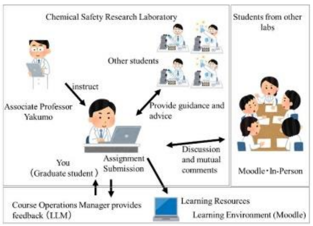


Figure 1. Learning flow of SCC with LLM-driven feedback.

Table 1. *LLM-Driven Feedback Prompt Components and ARCS-V Alignment*

Prompt Components	Description
Assignment input	Task content is inserted into the prompt (R)
Evaluation criteria	Task-specific criteria are provided to the LLM (C).
Role instruction	LLM acts as a university lab supervisor (A, R)
Motivational guidance	Satisfaction and Volition are explicitly requested (S, V).
Structured feedback	Overall, pass, strengths, advice, next expectations (S, C, V).
Concise output	Feedback limited to approximately 10 lines (S, V).

4. Results

All 14 feedback comments followed the structure specified in the prompt. In the overall comments, the LLM consistently acknowledged students' effort, and all assignments were judged as "Pass," which aligned with the human evaluator's judgment. Regarding the content of the feedback, for example, in Assignment 1 (accident case investigation), the LLM highlighted as strengths the clear linkage between accident causes and preventive measures, as well as the student's ability to summarize safety requirements in their own words. In the advice for improvement, the LLM pointed out that the student's statement "acetonitrile has low volatility" was inaccurate and encouraged checking physicochemical properties using reliable sources. In the expectations for the next assignment, the LLM suggested that focusing on hazard information and its connection to laboratory behavior in Assignment 2 would help the student grow as a safety-conscious researcher. These comments were generally appropriate and supportive of learner motivation. However, the remark about verifying physicochemical properties was somewhat misaligned with the course sequence, as such investigation is the focus of the following assignment. This indicates that while the feedback was largely valid, some comments could better account for task progression across assignments.

To further examine the pedagogical appropriateness of the generated feedback, a chemical safety specialist at a university conducted a formative evaluation using criteria such as accuracy, practicality, appropriateness as instructor feedback, contribution to sustaining learner motivation, and ability to stimulate a sense of accomplishment. The Subject Matter Expert (SME) judged the feedback to be appropriate in all categories. Importantly, the SME noted that the AI-generated feedback sometimes stimulated student satisfaction more effectively than human-written feedback. However, because all evaluated assignments met the passing criteria, the SME pointed out that the system's ability to address incorrect or unsafe responses remains untested.

5. Discussion and Conclusion

The results suggest that LLM-generated feedback can support sustained learning in chemical risk management. Prompts designed based on ARCS-V appeared to enhance learner motivation, particularly in terms of satisfaction and volition. However, this evaluation was limited to 14 submissions that met passing criteria; further validation with more diverse data, including incorrect or safety-compromised responses, is essential to ensure reliability. In addition, LLM-generated feedback may be prone to inaccuracies and an overly affirmative tone. To address these limitations while retaining the benefits of rapid feedback, a human–AI collaborative approach is recommended, combining AI-generated comments with peer review and instructor feedback. This study examines the feasibility of integrating motivational design principles into SCC-based learning rather than assessing the intrinsic performance of the LLM. The findings suggest that the effectiveness of LLM-generated feedback depends not only on model capability but also on how it is embedded within sound instructional design.

References

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

Katayama, K., Kita, T., Nakano, H., & Goda, Y. (2024). Design of an e-learning course to enhance autonomy of chemical risk management. Proceedings of the 45th JSET National Conference (Autumn 2024), 717–718.

Keller, J. M. (2008). First principles of motivation to learn and e3-learning. Distance Education, 29(2), 175–185.

Matelsky, J. K., Parodi, F., Liu, T., Lange, R. D., & Kording, K. P. (2023). A large language model-assisted education tool to provide feedback on open-ended responses. arXiv:2308.02439.

Mazzullo, E., & Bulut, O. (2024). Automated feedback generation for open-ended questions: Insights from fine-tuned LLMs. Proceedings of Machine Learning Research (FM-EduAssess at NeurIPS 2024 Workshop), 1–18.

Schank, R. C. (2007). The Story-Centered Curriculum. eLearn Magazine

Hsu, T. C., & Ogata, H. (Eds.). (2026). *Proceedings of the 10th International Conference on Artificial Intelligence, Computational Thinking and STEM Education* .Kumamoto: The Asia-Pacific Society for Computers in Education.

Elementary Students' Perceptions of Information Technology in Physical Computing Lessons across Different Learning Topics

Moe TANAKA [1*], Akinobu ANDO [2], Kazutoshi NUKADA [3] & Tatsuya HORITA [4]
[1,4] Graduate School of Teacher Education, Tokyo Gakugei University, Japan
[2] Hiroshima Institute of Technology, Japan
[3] e-Craft Inc., Japan
*moe.dv3vb.178@gmail.com

Abstract: *This study examined upper elementary students' perceptions of information technology through three physical computing lessons using different learning topics with sensors and servo motors. The students' free-response descriptions were categorized into ten types. The results indicated that "understanding of mechanisms" was most frequent in the Smart Home lesson, "functions of information technology" in the Dam system lesson, and "convenience and reduction of human effort" in the Smart Greenhouse lesson. These findings suggest differences in students' perceptions of information technology across learning topics.*

Keywords: Elementary education, physical computing, sensors

1. Introduction

In Japanese elementary programming education, students are expected to understand the mechanisms and characteristics of information technology (Ministry of Education, Culture, Sports, Science and Technology, 2020). Unfortunately, it has been pointed out that students' understanding remains superficial (Horita, 2019). To address this issue, several studies on physical computing have been conducted (Muratsu & Takahashi, 2023; Kuroda & Moriyama, 2020). However, few studies have examined how students' perceptions of information technology differ across learning topics. We assumed that students' learning experiences and perceptions of information technology might change depending on the topics used in physical computing lessons. Therefore, this study aimed to conduct physical computing lessons using three different approaches and analyze how elementary school students perceive information technology.

2. Methods

2.1. Learning Materials and Overview of the Three Physical Computing Lessons

We used the embot+ starter kit (e-Craft Inc.) as the learning material. This kit includes a microcontroller board called the "embot+ core" and cardboard pieces shaped like a character. This core can connect up to three sensors, three servo motors, two LEDs, one buzzer, and one speaker. Depending on their needs, users have the flexibility to use only the core or to adapt the cardboard components for their own designs.

We provide an overview of our three physical computing lessons. Table 1 summarizes the subjects, grades, goals, and learning activities for each lesson. Each session used a different topic, with different sensors, control programs, and algorithms. Figure 1 shows examples of the models from the three physical computing lessons.

Table 1. *An overview of the three curriculum units conducted*

Lesson 1 (Smart Home) :6th Grade Home Economics (Figure 1:a)	
Goals	Creation of a smart home system that closes curtains as light intensity increases, using a light sensor (LS), servo motors, LEDs, a buzzer (BZ), and a house model.
Learning Activities	①Creating programs for sensing ②Creating conditional logic to process sensor data ③Constructing a curtain mechanism that operates when a threshold is exceeded
Lesson 2 (Dam Flood Mitigation System) :5th Grade Period for Integrated Studies (Figure 1:b)	
Goals	Creation of a flood mitigation system that opens floodgates automatically as water levels rise, using a water sensor, servo motors, and a dam model.
Learning Activities	①Creating programs for sensing ②Creating conditional logic to process sensor data ③Constructing a floodgate mechanism that operates when a threshold is exceeded
Lesson 3 (Smart Greenhouse) :5th Grade Social Studies (Figure 1:c)	
Goals	Creation of a smart greenhouse system that automatically opens the plastic cover as temperature increases, using a temperature/humidity (T/H) sensor, servo motors, LEDs, a buzzer, and a greenhouse model.
Learning Activities	①Creating programs for sensing ②Creating conditional logic to process sensor data ③Constructing a greenhouse mechanism that operates when a threshold is exceeded

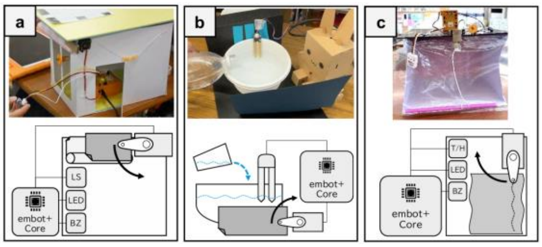


Figure 1 Model examples from the three physical computing lessons

a. Lesson 1 (Smart Home).b. Lesson 2 (Dam Flood Mitigation System) c.Lesson 3 (Smart Greenhouse)

2.2. Evaluation Method

To examine students' perceptions of information technology, students were asked to describe what they learned about sensors and computers through programming. In this study, to consistently investigate students' perceptions of information technology, students were asked to write what they had learned about sensors and computers through programming. In Lesson 1, responses from 16 sixth-grade students at a public elementary school were analyzed. In Lessons 2 and 3, responses from 18 fifth-grade students were analyzed. All participating students had prior experience using embot+ for four hours. The responses were categorized into ten types based on content similarity.

A. Functions of information technology, B. Application in society and daily life,

C. Understanding of mechanisms, D. Convenience and reduction of human effort,

E. Characteristics of information technology, F. Need for human involvement, G. Adjustment of threshold values, H. Difficulty of creating programs, I. Collaboration with friends, J. Others

3. Results

Figure 2 shows the classification results. The most frequent categories in Lesson 1, 2, and 3 were "C. Understanding of mechanisms" (36.8%), "A. Functions of information technology" (32.0%), and "D. Convenience and reduction of human effort" (23.2%), respectively. Differences in students' perceptions were observed across the lessons.

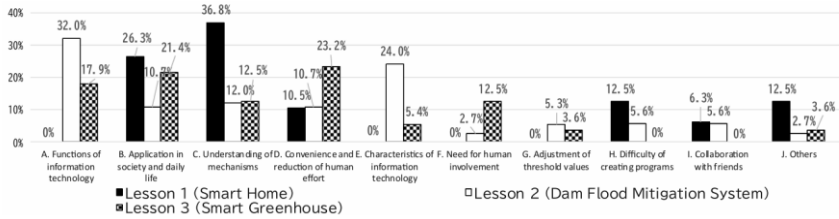


Figure 2 Understanding Sensors and Computers through Programming.

4. Discussion

In Lesson 1, C (Understanding of mechanisms) was the most frequent category. Reflections such as “it moves exactly as programmed” suggest that students recognized how the smart home operated. This result implies that engaging in mechanism construction may promote recognition of programming logic. In Lesson 2, A (Functions of information technology) was the most frequent category. Reflections such as “sensors can be controlled by a computer” indicate that students recognized the functional roles of sensors and computers. This finding suggests that developing sensing programs and applying conditional logic may have contributed to students’ recognition of these technological functions. In Lesson 3, D (Convenience and reduction of human effort) was the most frequent category. Reflections such as “it moves automatically thanks to the computer, even if no one is there” indicate that students recognized the benefits of automation. This finding suggests that constructing an automated greenhouse model may have contributed to students’ recognition of how automation can reduce human effort.

5. Conclusions

This study suggests that understanding different aspects of information technology may require programming experiences across multiple topics using sensors and servo motors. By implementing physical computing lessons with different approaches, the findings showed that students’ perceptions of information technology varied according to the learning topic.

Future research should further examine how these lessons relate to science, particularly physical concepts that support deeper understanding of the phenomena students experienced, as well as to mathematics for understanding and handling data. It is also necessary to examine the role of such lessons in the school curriculum.

References

e-Craft Inc. (n.d.). embot+. https://www.embot.jp/

Horita, T. (2019). What school education should be like in the AI era. Science Council of Japan Forum, 24(2), 2_56-2_65. (in Japanese)

Kuroda, M., & Moriyama, J. (2020). Practice and effects of programming education to promote understanding of the social roles of information technology in elementary school social studies. Journal of Educational Information, 36(2), 75-86. (in Japanese)

Ministry of Education, Culture, Sports, Science and Technology. (2020). Guide to Programming Education in Elementary Schools (3rd ed.). (in Japanese)

Muratsu, K., & Takahashi, O. (2023). Lesson design to deepen understanding of the usefulness of digital sensors in science: A preliminary study on "Plant nutrients and water pathways" for 6th-grade elementary school students. Proceedings of the Annual Meeting of Japan Society for Science Education, 47, 751-754. (in Japanese)

Hsu, T. C., & Ogata, H. (Eds.). (2026). *Proceedings of the 10th International Conference on Artificial Intelligence, Computational Thinking and STEM Education* .Kumamoto: The Asia-Pacific Society for Computers in Education.

LLM-Based Question Generation System for Lecture Videos with Question-Quality Evaluation

Yuya YOSHIDA [1*], Koichi YOSHIZAKI [2], Minoru IKEBE [3] & Hikaru ICHISE [4]
[1,2,3,4] Oita University, Japan
*v25y2018@oita-u.ac.jp

Abstract: *Although assessment quizzes enhance learner understanding, their manual creation requires careful review of lecture content and verification of answer correctness, placing a significant burden on instructors. This study developed an LLM-based system to automatically generate and evaluate questions from lecture videos. In this system, on-screen text and spoken content in the video are extracted using Optical Character Recognition (OCR) and Automated Speech Recognition (ASR), serving as contextual information for the LLM. To reduce the extraction cost, OCR is applied only to the necessary frames detected by slide-change analysis. To improve ASR for specialized terms, the system generates a proper-noun list using an LLM and appends it to the ASR prompts. Instructors' evaluations on a five-point scale showed that questions generated by gpt-oss and qwen3-vl achieved mean accuracy scores of 4.0 or higher. However, the system's automatic evaluation showed only weak agreement with instructor ratings, suggesting current limitations in LLM-based question evaluation.*

Keywords: automatic question generation, automatic question evaluation, lecture video, LLM, ASR, OCR

1. Introduction

In educational institutions, assessment quizzes are used to promote learners' understanding. However, manually creating these quizzes requires instructors to carefully review lecture content, which imposes a significant workload. Molina et al. (2024) proposed an automatic question generation method using lecture slides and reported high relevance between the generated questions and the slide content. Nonetheless, their approach remains limited by the need to manually extract key points from slides. Our prior work (Yoshida et al., 2025) examined automatic question generation from Japanese lecture videos using OCR, ASR, and LLM to eliminate manual context extraction. However, this approach faced two problems. First, extracting information from lecture videos is time-consuming. Second, OCR/ASR recognition errors often propagate into the generated questions. Furthermore, evaluating the quality of the questions relied on manual scoring. While frameworks like G-EVAL (Liu et al., 2023) have shown the potential of Chain-of-Thought (CoT) for evaluating natural language generation outputs, achieving a Spearman correlation of 0.51 with human judgment, their application to Japanese pedagogical contexts remains under-explored. Consequently, this study pursues two objectives: First, the video information extraction and question generation pipeline is redesigned to improve both computational efficiency and question quality. Second, inspired by G-EVAL, an automatic evaluation method is introduced to enable question quality verification without relying on manual scoring.

2. Proposed Method

Our proposed system consists of three stages: information extraction (OCR/ASR), question generation, and automatic question evaluation. Experiments were conducted on a MacBook Pro with an Apple M1 Pro chip and 32 GB of memory. The details of each stage are described in the following Sections 2.1–2.2.

2.1. Information Extraction (OCR/ASR)

This stage extracts textual information from lecture videos using OCR and ASR. First, slide-change points are detected with ffprobe by identifying frames with large pixel differences. Next, GLM-OCR is applied to representative frames at these points to recognize on-screen slide text, and the extracted text is stored along with its corresponding timestamp range. In parallel, the audio track is transcribed using mlx-Whisper. To improve recognition of domain-specific proper nouns, the system generates a proper-noun list from the OCR results and provides it to mlx-Whisper as an initial prompt. By propagating this initial prompt across all subsequent segments, the system ensures consistent terminology throughout the entire transcription. Finally, the recognized text and its timestamps are stored in a database.

2.2. Question Generation/ Question Evaluation

Questions are generated from the extracted OCR/ASR context using an LLM. To reduce misspellings and inconsistent surface forms, the proper-noun list used for ASR is also incorporated into the question-generation prompt. Question generation is conducted using three models: gpt-oss:20b (8-bit quantization), qwen3-vl:30b (4-bit quantization), and Llama 3.1:8b (8-bit quantization). The generated question–answer pairs are stored in a database. To evaluate question quality automatically, we implemented an automatic evaluation module inspired by G-EVAL. This stage has 3 steps. First, instructors write evaluation criteria and rubrics. Second, the system calls the LLM to generate a Chain-of-Thought (CoT), which is a detailed evaluation method. Finally, questions are evaluated by inputting the question text, answer text, CoT, and corresponding context for each question into the LLM. Two LLMs (gpt-oss:20b (8-bit quantization), qwen3-vl:30b (4-bit quantization)) are used for evaluation.

3. Experimental Results

The proposed system was used to automatically generate questions for two courses: “Introduction to Oita University” and “Computer Network.” The course “Introduction to Oita University” contains numerous proper nouns, including building names and place names. “Computer Network” is a lecture course in computer science.

3.1. Processing time for OCR

In previous studies, frequent OCR processing was a bottleneck. In contrast, this system minimized the number of OCR operations by using ffprobe to perform OCR only in areas with pixel value changes. As a result, the processing time for this method was reduced to an average of 5 minutes and 23 seconds across five runs using the same 80-minute video.

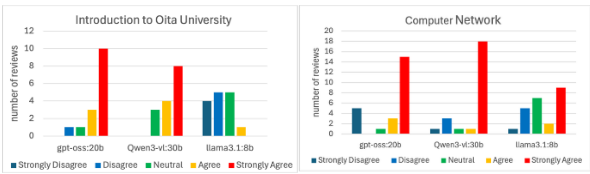


Figure 1. Results of instructor evaluation for automatically generated questions

3.2. Question generation result

Figure 1 shows the instructor’s accuracy rating. For “Introduction to Oita University,” the mean scores were 4.47 for gpt-oss, 4.33 for Qwen3-VL, and 2.20 for Llama3.1. t-tests indicated significant differences between gpt-oss and Llama3.1 and between Qwen3-VL and Llama3.1, whereas the difference between gpt-oss and Qwen3-VL was not significant, suggesting similar accuracy. For “Computer Network,” the mean scores were 3.96 for gpt-oss, 4.33 for Qwen3-VL, and 3.54 for Llama3.1. No significant differences were observed between gpt-oss and Qwen3-VL or between gpt-oss and Llama3.1. However, a significant difference was observed between Qwen3-VL and Llama3.1, with Qwen3-VL showing higher accuracy.

3.3. Correlation Between Human Evaluation and LLM Evaluation

Table 1. *Spearman's Rank Correlation (ρ) Between Human and LLM Evaluations*

	Evaluation LLM	
Video name	gpt-oss:20b	qwen3-vl:30b
Introduction to Oita University	0.27	0.51
Computer Network	0.32	0.36

Spearman's rank correlation coefficient (ρ) was computed between the instructor's accuracy ratings and the LLM evaluation scores (Table 1). For "Introduction to Oita University," gpt-oss showed a weak correlation (ρ = 0.27), whereas qwen3-vl achieved a moderate correlation (ρ = 0.51). For "Computer Network," correlations were modest for both evaluators (gpt-oss: ρ = 0.32; qwen3-vl: ρ = 0.36). In addition to correlation analysis, the mean difference between instructor ratings and LLM evaluation scores was examined to assess systematic scoring tendencies. For "Introduction to Oita University," the average difference was 0.522 for gpt-oss and 1.170 for qwen3-vl. For "Computer Network," the average difference was less than 0.001 for gpt-oss and 0.648 for qwen3-vl. These results indicate that both models tended to assign higher scores than the instructor, suggesting a leniency bias in LLM-based evaluation. Notably, the bias was particularly pronounced for qwen3-vl in "Introduction to Oita University". To investigate these scoring discrepancies, we manually corrected OCR omissions caused by animations and restricted the input context range specifically to the parts relevant to the generated questions. This additional experiment improved agreement with human ratings in half of the cases, indicating that context optimization is critical for enhancing evaluation accuracy.

4. Conclusion

This paper proposed a framework that directly generates questions and evaluates question quality from lecture videos. In instructor evaluations, questions generated by the gpt-oss:20b and Qwen3-VL:30b models achieved accuracy scores of 4.0 or higher. However, correlations between human and LLM-based accuracy evaluations remained weak to moderate. Future work will focus on automating the context optimization process—specifically, extracting video segments relevant to the generated questions—which our manual analysis identified as critical for reducing scoring discrepancies. Additionally, we plan to mitigate LLM's leniency bias and investigate the practical applications of the generated questions.

References

Molina, I. L., Švábenský, V., Minematsu, T., Chen, L., Okubo, F., & Shimada, A. (2024). Comparison of large language models for generating contextually relevant questions. https://doi.org/10.1007/978-3-031-72312-4_18

Yoshida, Y., Yoshizaki, K., Ikebe, M., & Ichise, H. (2025). Automatic generation of review tests based on lecture videos using LLM. https://doi.org/10.11527/jceeek.2025.0_398

Liu, Y., Iter, D., Xu, Y., Wang, S., Xu, R., & Zhu, C. (2023). G-Eval: NLG evaluation using GPT-4 with better human alignment. https://doi.org/10.18653/v1/2023.emnlp-main.153

Hsu, T. C., & Ogata, H. (Eds.). (2026). *Proceedings of the 10th International Conference on Artificial Intelligence, Computational Thinking and STEM Education* .Kumamoto: The Asia-Pacific Society for Computers in Education.

Design and Development of an Interactive Chatbot for Promoting Readiness for Class in Middle School Science

Maria Suzune OBA[1*], Hisashi HATAKEYAMA[2], Masao MUROTA[3]
[1,2,3] Institute of Science Tokyo, Japan
*oba.m.1203@m.isct.ac.jp

Abstract: *Students with average or below-average academic achievement often enter class without sufficient readiness for class, resulting in reduced willingness to participate. This lack of readiness arises when insufficient prior knowledge increases cognitive load, leading to heightened anxiety and decreased engagement. To address this issue, this study designs a dialogue-based review chatbot that supports learners' readiness for class, defined as a psychological and cognitive preparatory state involving reduced anxiety, perceived ease of engagement, and partial recall of prior content. The chatbot adopts a four stage dialogue structure - empathy, retrieval, guided explanation, and successful completion - grounded in Bandura's four sources of self-efficacy. A large language model extracts the structural elements of science worksheets, and a state-transition controller generates unit independent dialogue flows. The proposed system does not aim to measure knowledge, but to facilitate a psychologically safe pre class experience that promotes learners' expectations of success.*

Keywords: interactive review chatbot, peer-like conversational agent, self-efficacy

1. Introduction

Students with average to below-average academic achievement often struggle to engage in classroom activities due to insufficient readiness for class. Insufficient prior knowledge increases cognitive load, which is associated with lower engagement (Dong et al., 2020). Prompting learners to recall prior content may foster self-efficacy, which influences engagement and persistence (Bandura, 1997).

Retrieval practice has been shown to improve performance and reduce anxiety (Agarwal et al., 2012). However, when recall activities are expose learner's inability, they may increase anxiety and discourage participation. These considerations highlight the importance of designing psychologically safe pre-class retrieval activities that support learners' readiness for class.

1.1. Related Work and Research Gap

Generative AI–based chatbots have been used as learning support tools that enhance engagement and self-efficacy when providing process-oriented guidance rather than direct answers (Au et al., 2025; Lee et al., 2024). However, prior research on dialogue-based learning support has largely focused on engagement and learning outcomes during or after instruction, with comparatively less attention given to learners' readiness for class.

Addressing this gap, the present study designs a short pre-class dialogue intervention to support learners' readiness and develops a reusable dialogue-generation chatbot based on structural extraction. In this study, readiness for class refers to a psychological and cognitive preparatory state characterised by (a) reduced anxiety about the class, (b) perceived ease of engaging with the learning content, and (c) a sense of being able to recall at least part of the content learned beforehand.

1.2. Objectives and Contributions

The present study designs a chatbot that supports learners' readiness for class through brief pre-class dialogues, focusing on preparatory state rather than academic performance. It adopts a staged structure of empathy, retrieval, guided explanation, and successful completion to alleviate performance-related stress, particularly among learners with average to below-average achievement.

This paper makes three contributions. First, it explicitly positions learners' psychological and cognitive readiness for classroom participation as a design target. Second, it applies Bandura's (1997) self-efficacy framework into a staged dialogic design model for pre-class support. Third, it presents a reusable dialogue-generation structure based on science experiment worksheet templates.

2. System Overview and Design

The proposed chatbot is designed not to assess learners' knowledge, but to support readiness for class through pre-class dialogue. It focuses on learners' preparatory state prior to classroom participation.

2.1. Design Principles and Agent Functions

The design is grounded in Bandura's (1997) four sources of self-efficacy: enactive mastery, vicarious experience, verbal persuasion, and physiological and affective states. These are embedded in the dialogue structure to support learners' readiness. Table 1 summarizes the corresponding design principles and implementations.

Table 1. *Design and Implementation Based on Bandura's Self-Efficacy Framework*

Source of Self-Efficacy	**Design Intent**	**Specific Behaviours**	**Implementation Elements**
Enacti mastery	Stepwise success	Affirmative feedback; retry-to-success	LangGraph state control + staged hints + MCQ LLM
Vicarious Experience	Peer demonstrates desirable strategies	Modelling learning strategies (e.g., textbook cross-checking)	Peer persona prompt + UI guidance
Verbal Persuasion	Success expectancy support	Non-judgmental feedback	Empathetic, non-judgmental feedback control
Physiological & Affective States	Anxiety reduction	Error-label suppression + empathy	Correctness-label suppression + tone regulation

Each source is operationalized through specific interaction strategies, such as stepwise feedback, peer modelling, non-judgmental responses, and anxiety-reducing dialogue.

2.2. System Architecture

This system is operationalized through specific LLM functions and system-level control mechanisms (Figure 1). The system consists of four layers: worksheet input, LLM-based structure extraction, LangGraph state control, and utterance generation with UI presentation. LangGraph manages state transitions and structural decisions, while the LLM is used solely for language generation, enabling unit-independent dialogue flows.

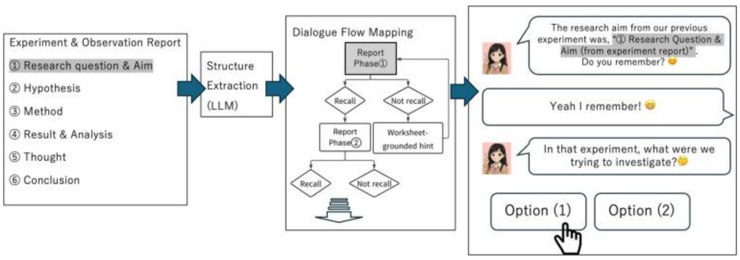


Figure 1. Framework for Generating Worksheet-Grounded Pre-class Recall Dialogue

Figure 2 illustrates the dialogue interface. The left panel displays instructional inputs, and the right panel presents peer-like interactions delivered in small, multiple-choice steps to prompt retrieval and guide understanding.

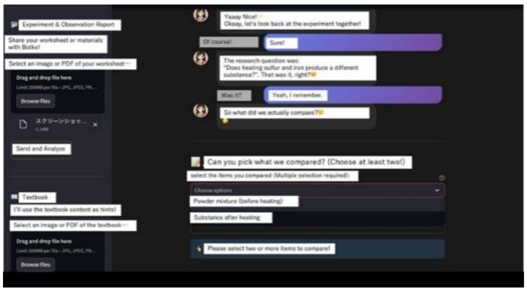


Figure 2 Example of the Peer-Based Pre-class Retrieval Dialogue

3. Future Work

Future work will evaluate the chatbot's impact on learners' readiness and classroom engagement and examine the generalizability of the worksheet-based dialogue structure across multiple units. As the study involves minors, it will be conducted only after review and approval by the university's institutional ethics committee.

References

Agarwal, P. K., Bain, P. M., & Chamberlain, R. W. (2012). The Value of Applied Research: Retrieval Practice Improves Classroom Learning and Recommendations from a Teacher, a Principal, and a Scientist. Educational Psychology Review, 24(3), 437–448.

Au, C. K., Yoo, S., & Chiu, T. K. F. (2025). Student Engagement and Teacher Perceived Support in STEAM Education Using Generative AI: A Systematic Review and Direction for Future Research. School Science and Mathematics.

Bandura, A. (1997). Self-efficacy: The exercise of control. Macmillan.

Dong, A., Jong, M. S. Y., & King, R. B. (2020). How Does Prior Knowledge Influence Learning Engagement? The Mediating Roles of Cognitive Load and Help-Seeking. Frontiers in Psychology, 11, 591203.

Lee, H. Y., Chen, P. H., Wang, W. S., Huang, Y. M., & Wu, T. T. (2024). Empowering ChatGPT with guidance mechanism in blended learning: effect of self-regulated learning, higher-order thinking skills, and knowledge construction. International Journal of Educational Technology in Higher Education, 21(1).

2026

AI -CTE-STEM

Website URL:

https://eds.let.media.kyoto-u.ac.jp/TBICS2026/

Email:

AI-CTE-STEM2026@easychair.org

APSCE

熊本大学
Kumamoto University

KYOTO UNIVERSITY KYOTO JAPAN FOUNDED 1897

師大