Learning Outcomes, Cognitive Load, and Interaction Patterns in Generative AI-Supported Elementary Mathematics Learning

Bor-Chen Kuo[1], Pei-Chen Wu[1,*], Chen-Huei Liao[1] and Cheng-Hsuan Li[1]

[1]*National Taichung University of Education, Taichung, Taiwan*

Abstract

This study presents a preliminary student evaluation of Joyful Learning TALPer, an LLM-based scaffolding tutor. The system is designed to provide dialogic learning support through Explicit Instruction and Socratic questioning, with tutor strategies including prompting, scaffolding, modeling, feedback, sequencing, summarizing, and affective support. The present study examined its use in an elementary mathematics learning activity. A total of 21 fifth-grade students used Joyful Learning TALPer to learn a unit on multiples and common multiples. Pre- and post-tests were used to examine changes in mathematics performance. Learning satisfaction and cognitive load were measured after the activity. In addition, Epistemic Network Analysis (ENA) was conducted to examine patterns of connections among the tutor's instructional strategies during student–AI tutor mathematics dialogues. Results showed significant improvement in students' mathematics test performance. After the learning activity, students reported positive learning satisfaction and a moderate level of cognitive load. The cognitive load results indicated that students perceived the learning content, task demands, instructional presentation, and explanation format as manageable. The ENA results revealed identifiable connections among these strategies, showing that scaffolding, feedback, prompting, sequencing, and affective support were closely integrated during student–AI tutor interactions. These findings provide preliminary evidence for the feasibility of applying Joyful Learning TALPer in elementary mathematics learning and offer a basis for future studies involving students with learning difficulties or low academic achievement.

Keywords

Large Language Models, Scaffolding Tutor, Elementary Mathematics, Cognitive Load, Epistemic Network Analysis

1. Introduction

With the rapid development of generative artificial intelligence and large language models (LLMs), conversational AI systems have increasingly been applied in educational contexts. Such systems can respond to student input in real time, provide individualized explanations, and offer support that may be difficult to provide immediately in ordinary classrooms. Prior studies have discussed the potential of LLMs for educational assistance, intelligent tutoring, and feedback generation [5, 7, 10].

Pedagogically appropriate guidance requires more than correct answers or complete explanations. In instructional dialogue, students often need support in understanding task requirements, following step-by-step reasoning, revising incorrect responses, and maintaining engagement. Students with special educational needs, learning difficulties, or low achievement often require differentiated instructional strategies, including instructional scaffolding and flexible approaches [3, 11].

Cognitive load is also important when generative AI chatbots are integrated into learning activities. Chatbots may help students manage cognitive demands through simplified explanations and step-by-step guidance, but students must also follow the dialogue and interpret AI-generated responses. Prior K–12 research suggests that chatbot design may shape interaction patterns and cognitive load management [6].

HAI-Agency: Workshop on Orchestrating Human and AI Agency for Proactive and Reflective Learning, co-located with the 27th International Conference on Artificial Intelligence in Education (AIED 2026), Seoul, Korea, June 27 - July 3, 2026.

*Corresponding author.

kbc@mail.ntcu.edu.tw (B. Kuo); pedropcwu@gmail.com (P. Wu); chenhueiliao@gmail.com (C. Liao); ChengHsuanLi@gmail.com (C. Li)

However, several gaps remain in the current literature. Although educational chatbots have been studied across purposes, educational levels, and interaction mechanisms [12], relatively few studies explain how pedagogical principles are translated into concrete designs for LLM-based tutors [10]. In addition, although recent K–12 research has examined students' perceptions, interaction patterns, and learning support with generative AI chatbots [6], evidence remains limited in elementary mathematics, especially for cognitive load and post-activity perceptions. Furthermore, tutor moves during student–AI dialogues are often not examined in detail, making it difficult to determine whether the system functions as a pedagogically guided tutor or a general-purpose chatbot [10, 12].

In response, this study examines how Joyful Learning TALPer can be implemented as a pedagogically guided LLM-based tutor in an elementary mathematics learning activity. Joyful Learning TALPer is an LLM-based scaffolding tutor grounded in Explicit Instruction [1], Socratic questioning, and differentiated support for diverse learners [3, 11], utilizing simplified language, step-by-step guidance, and affective encouragement. The system provides structured dialogic support through prompting, scaffolding, feedback, sequencing, summarizing, modeling, and affective support.

In the present study, Joyful Learning TALPer was applied to an elementary mathematics unit on multiples and common multiples. As an initial student evaluation, this study examined fifth-grade students' mathematics performance, learning satisfaction, perceived cognitive load, and student–AI tutor interaction patterns. ENA was used to examine patterns of connections among tutor instructional strategies.

Accordingly, this study addressed the following research questions:

RQ1. Did students' mathematics test performance improve after using Joyful Learning TALPer in a mathematics learning unit?

RQ2. What levels of learning satisfaction and cognitive load did students report after using Joyful Learning TALPer?

RQ3. What patterns of connections among the tutor's instructional strategies emerged in student–AI tutor mathematics interactions?

2. System Design

Joyful Learning TALPer is an LLM-based scaffolding tutor. The system uses a structured prompt framework grounded in Explicit Instruction [1] and Socratic questioning, guiding it to focus on critical content, sequence skills logically, decompose complex tasks, provide guided practice, and offer immediate feedback.

In the present implementation, Joyful Learning TALPer was used in a fifth-grade mathematics unit on multiples and common multiples. The tutor model used in the learning activity was Gemini 2.5 Flash, and students interacted with the tutor through the Joyful Learning TALPer chat interface.

The tutoring flow was designed to be simple and consistent. The tutor was instructed to first check the student's current idea and then provide a suitable support move, such as a guiding question, hint, feedback, or short explanation. When a student gave an incomplete or incorrect answer, the tutor was instructed to give a simpler prompt or example first, and to avoid giving the final answer too early.

This design differs from a general-purpose chatbot in three ways. First, the prompt framework limited the tutor to the target mathematics unit and required stepwise guidance. Second, the tutor was asked to avoid long explanations and direct answer-giving when a hint or question was more suitable. Third, the system included affective support, such as short encouragement or reassurance, so that feedback was not only corrective but also supportive.

The tutor prompt specified the tutor role, interaction rules, instructional constraints, and other settings. Appendix A provides a representative shortened excerpt showing the role setting and selected interaction rules.

Tutor responses were coded according to their pedagogical functions. The coding framework included seven tutor instructional strategies: prompting (T.PROMPT), scaffolding (T.SCAFF), modeling (T.MODEL), feedback (T.FEED), sequencing (T.SEQ), summarizing (T.SUM), and affective support (T.AFF).

For example, T.PROMPT referred to questions or prompts used to elicit student thinking, such as "What is the first step we should try?", and was aligned with Explicit Instruction Principles P1 and P11. T.SCAFF referred to hints, partial guidance, or step cues, such as "You can first list the multiples of 4, and then list the multiples of 6," and was aligned with P3 and P10. T.FEED referred to immediate affirmative, corrective, or next-step feedback, such as "That is a good start, but let's check whether 12 appears in both lists," and was aligned with P13 and P12. The coding framework was used to examine how Joyful Learning TALPer translated Explicit Instruction principles into dialogic tutoring behaviors during mathematics learning interactions.

3. Method

3.1. Participants and Learning Unit

Participants were 21 fifth-grade students from a general elementary classroom. The learning unit was multiples and common multiples. The Joyful Learning TALPer learning activity lasted approximately 40 minutes, excluding the pre- and post-tests. Students completed a mathematics test before and after using Joyful Learning TALPer. They also completed learning satisfaction and cognitive load questionnaires after the learning activity.

3.2. Measurement Instruments

Students' mathematics performance was measured using pre- and post-tests on multiples and common multiples. Each test included eight multiple-choice items drawn from an expert-reviewed digital learning platform. The pre- and post-tests used parallel items covering the same concepts. Test scores were converted into percentage scores. Learning satisfaction was measured after the learning activity using a nine-item, 5-point Likert scale adapted from Chu et al. [2]. Cognitive load was also measured after the learning activity using an eight-item, 7-point Likert scale from Hwang et al. [4]. The cognitive load measure included two subscales: mental load and mental effort. Mental load referred to students' perceived difficulty, challenge, pressure, frustration, and time demand in the learning activity. Mental effort referred to students' perceived effort associated with the instructional format, presentation, and explanation of the learning content.

3.3. Data Analysis

Paired-samples t-tests were used to examine pre-post changes in mathematics test performance, with the direction of change defined as posttest minus pretest. Effect size was calculated as paired Cohen's d. Learning satisfaction and cognitive load were analyzed using descriptive statistics because these measures were administered only after the learning activity.

Epistemic Network Analysis (ENA) was used to examine patterns of co-occurrence among tutor instructional strategies in student–AI tutor dialogues. Only tutor turns were included in the ENA analysis. Each tutor response was coded according to its pedagogical function, and one response could receive one or more codes. Co-occurrences were modeled using a moving window of four coded tutor responses; therefore, nodes represented tutor instructional strategies, and edges represented the co-occurrence strength among strategies within the window [9]. The analysis was conducted using the ENA Web Tool (Version 1.9.0) [8].

For the coding process, the unit of analysis was one tutor response. The final analysis included the 21 students who completed both the pretest and posttest. Each student dialogue contained approximately 30 to 40 tutor turns. Each tutor response could receive one or more strategy codes because a single response could, for example, provide feedback while also providing affective support.

To ensure the reliability of the ENA coding, two human coders independently coded all tutor responses, and their inter-rater reliability reached Cohen's kappa = .847, indicating substantial agreement. Coding disagreements were discussed with reference to the codebook, and the agreed codes were used to prepare the final ENA dataset.

4. Results

4.1. Mathematics Performance

A paired-samples t-test was conducted to examine students' changes in mathematics performance before and after using Joyful Learning TALPer. Table 1 presents the descriptive and inferential results.

Students' mathematics test scores were significantly higher after the Joyful Learning TALPer activity, with an effect size approaching the large range.

Table 1: Pre-post comparison of mathematics performance

Variable	Pretest M (SD)	Posttest M (SD)	Mean difference	t	p	d
Mathematics test	28.81 (32.74)	60.38 (28.74)	31.57	3.535 **	.002	0.77

Note. Mean difference = posttest - pretest. ** p < .01.

4.2. Satisfaction and Cognitive Load

Descriptive statistics indicated that students reported positive satisfaction with the learning activity and a moderate level of cognitive load. Table 2 presents the item-level descriptive results.

The learning satisfaction mean was approximately 4.03 out of 5, suggesting that students generally viewed the activity positively. The cognitive load mean was approximately 3.86 out of 7, indicating a moderate level of perceived cognitive load after the activity. Because cognitive load was measured only after the activity, these results should be treated as descriptive post-activity perceptions.

Table 2: Satisfaction and cognitive load descriptive statistics

Variable	N	Item mean	Item-level SD
Learning satisfaction	21	4.03 / 5	0.96
Cognitive load total	21	3.86 / 7	1.91
Mental load	21	3.85 / 7	1.94
Mental effort	21	3.87 / 7	1.94

Note. Scores are reported as item-level means and standard deviations.

4.3. Interaction Patterns

ENA was conducted to examine patterns of connections among tutor instructional strategies during student–AI tutor mathematics interactions. As shown in Fig. 1, the network revealed identifiable tutor interaction patterns. T.SCAFF, T.FEED, T.PROMPT, T.AFF, and T.SEQ formed the main network, indicating that scaffolding, feedback, prompting, affective support, and sequencing frequently co-occurred during students' problem-solving dialogues.

T.AFF appeared visually prominent and formed visible connections with multiple instructional strategies, including T.PROMPT, T.FEED, T.SCAFF, and T.SEQ. This pattern suggests that affective support was embedded within the broader instructional process. T.MODEL and T.SUM were positioned outside the main network, indicating that modeling and summarizing were less central in the overall interaction patterns.

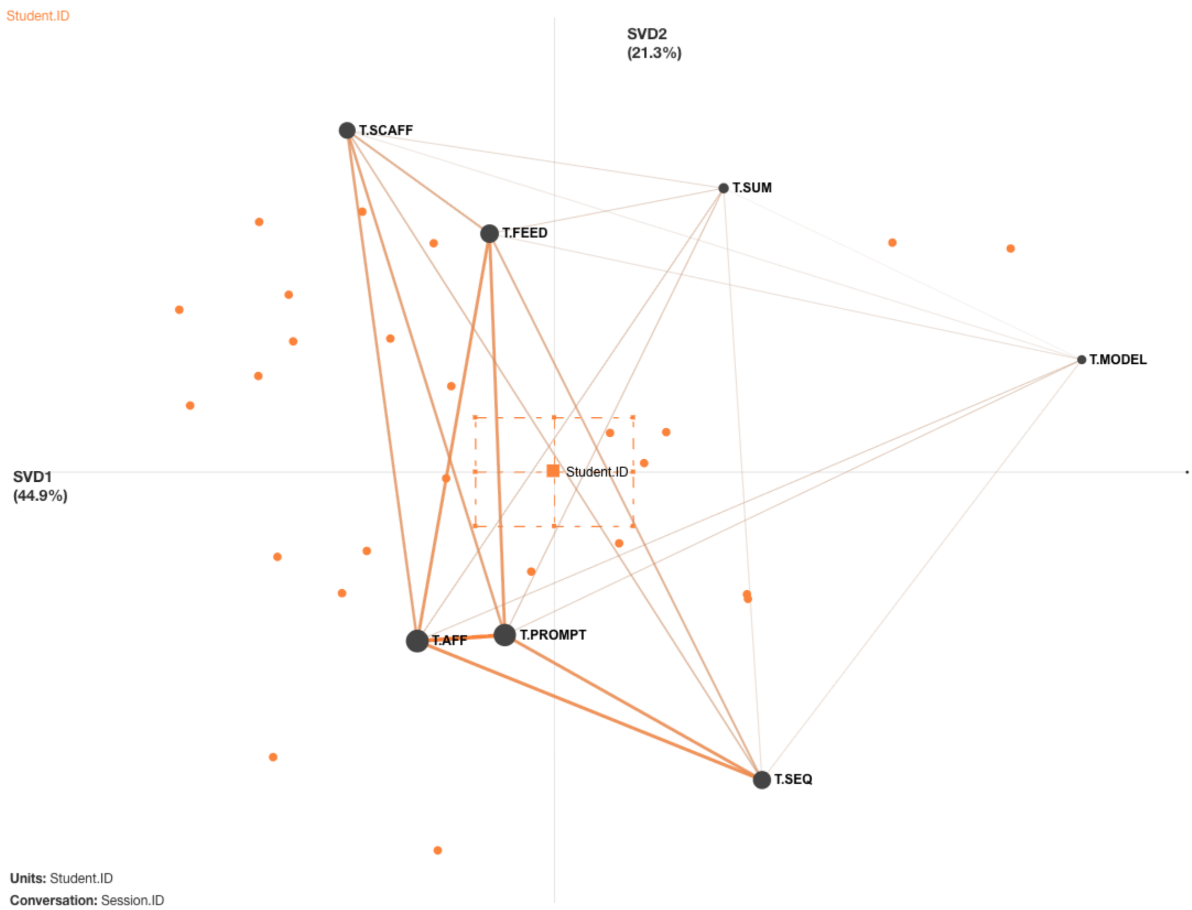

Figure 1: Tutor ENA network from student–AI tutor mathematics interactions. Nodes represent tutor instructional strategies, and edges represent co-occurrence strength among strategies within student–AI tutor dialogue sessions. Thicker edges indicate stronger co-occurrence patterns.

5. Discussion

Regarding RQ1, students' mathematics test scores were significantly higher after the learning activity. This finding suggests better performance on the unit after using Joyful Learning TALPer. However, because the study did not include a control group, the improvement should not be interpreted as causal evidence of system effectiveness.

Regarding RQ2, students reported positive satisfaction and a moderate level of cognitive load after the activity. These results suggest that students accepted the activity well and perceived the content and presentation as manageable. However, because cognitive load was measured only after the activity, the result does not show a reduction in cognitive load.

Regarding RQ3, the ENA results showed that scaffolding, feedback, prompting, sequencing, and affective support formed the main interaction pattern in the student–AI tutor dialogues. This suggests that TALPer combined instructional guidance with supportive responses during problem solving. Modeling and summarizing were less central, suggesting that the tutor mainly supported students through hints, feedback, and short prompts rather than long worked examples or end-of-task summaries. This may reflect the current emphasis on Socratic questioning and stepwise scaffolding, with modeling used more selectively when students could not progress through hints or feedback alone.

Overall, the findings provide preliminary support for applying Joyful Learning TALPer as a scaffolding tutor in elementary mathematics and for refining the system before extending it to students with learning difficulties or low academic achievement.

From a human–AI agency perspective, the tutor-side ENA results show how scaffolding, feedback, prompting, sequencing, and affective support were organized within TALPer's tutor-side responses.

Because the analysis focused only on tutor turns, it cannot directly explain students' learner agency or reflective learning. Future work will jointly analyze tutor moves and student responses to examine how proactive AI scaffolding shapes students' participation, reflection, and decision-making. In practice, TALPer should be viewed as a support tool rather than a replacement for the teacher, and the dialogue records may help teachers review students' support needs and plan whole-class or individual guidance.

6. Conclusion

This study conducted a preliminary evaluation of Joyful Learning TALPer in an elementary mathematics context. Students showed significantly higher mathematics test scores after using the system and reported positive satisfaction and moderate cognitive load. The ENA results showed tutor interaction patterns centered on scaffolding, feedback, prompting, sequencing, and affective support, suggesting that the system enacted its intended instructional strategies during student–AI tutor interactions.

This study has several limitations. First, it did not include a control group, and participants were from one general elementary class; therefore, the pre-post improvement should be interpreted as preliminary rather than causal evidence. Second, satisfaction and cognitive load were measured only after the activity and should be interpreted as post-activity perceptions. Third, although coding reliability reached substantial agreement, the ENA findings were based on coded tutor responses and may be influenced by the coding framework and final agreed codes. Future studies should include controlled designs, larger samples, students with learning difficulties or low academic achievement, and additional coding checks across learning units.

7. Acknowledgements

The authors gratefully acknowledge "The Digital Learning Enhancement Plan for Grades 1–12 Students" and the "Taiwan Adaptive Learning Platform," supported by the Ministry of Education, Taiwan, for providing the data used in this study. This work was also supported by the National Science and Technology Council, Taiwan, under Grant No. NSTC 112-2410-H-142-017-MY3.

8. Declaration on Generative AI

During the preparation of this work, the authors used ChatGPT (OpenAI) for language editing and improving the clarity of the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] A. L. Archer, C. A. Hughes, Explicit Instruction: Effective and Efficient Teaching, Guilford Publications, New York, NY, 2010.

[2] H. C. Chu, G. J. Hwang, C. C. Tsai, J. C. R. Tseng, A two-tier test approach to developing location-aware mobile learning systems for natural science courses, Computers & Education 55(4) (2010) 1618–1627. doi:10.1016/j.compedu.2010.07.004.

[3] T. S. Hasselbring, C. H. W. Glaser, Use of computer technology to help students with special needs, The Future of Children 10 (2000) 102–122. doi:10.2307/1602691.

[4] G. J. Hwang, L. H. Yang, S. Y. Wang, A concept map-embedded educational computer game for improving students' learning performance in natural science courses, Computers & Education 69 (2013) 121–130. doi:10.1016/j.compedu.2013.07.008.

[5] E. Kasneci, K. Sessler, S. Küchemann, M. Bannert, D. Dementieva, F. Fischer, U. Gasser, G. Groh, S. Günnemann, E. Hüllermeier, S. Krusche, G. Kutyniok, T. Michaeli, C. Nerdel, J. Pfeffer, O. Poquet, M. Sailer, A. Schmidt, T. Seidel, M. Stadler, J. Weller, J. Kuhn, G. Kasneci, ChatGPT for good? On

opportunities and challenges of large language models for education, Learning and Individual Differences 103 (2023) 102274. doi:10.1016/j.lindif.2023.102274.
[6] M. Klar, Using ChatGPT is easy, using it effectively is tough? A mixed methods study on K–12 students' perceptions, interaction patterns, and support for learning with generative AI chatbots, Smart Learning Environments 12 (2025) 32. doi:10.1186/s40561-025-00385-2.
[7] C. K. Lo, What is the impact of ChatGPT on education? A rapid review of the literature, Education Sciences 13(4) (2023) 410. doi:10.3390/educsci13040410.
[8] C. L. Marquart, C. Hinojosa, Z. Swiecki, B. Eagan, D. W. Shaffer, Epistemic Network Analysis Web Tool (version 1.9.0), 2021. URL: http://app.epistemicnetwork.org.
[9] D. W. Shaffer, W. Collier, A. R. Ruis, A tutorial on epistemic network analysis: Analyzing the structure of connections in cognitive, social, and interaction data, Journal of Learning Analytics 3(3) (2016) 9–45. doi:10.18608/jla.2016.33.3.
[10] J. Stamper, R. Xiao, X. Hou, Enhancing LLM-based feedback: Insights from intelligent tutoring systems and the learning sciences, in: A. M. Olney, I. A. Chounta, Z. Liu, O. C. Santos, I. I. Bittencourt (Eds.), Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners, Doctoral Consortium and Blue Sky, volume 2150 of Communications in Computer and Information Science, Springer, Cham, 2024, pp. 32–43. doi:10.1007/978-3-031-64315-6_3.
[11] C. A. Tomlinson, How to Differentiate Instruction in Academically Diverse Classrooms, 3rd ed., ASCD, Alexandria, VA, 2017.
[12] J. F. Urquiza-Yllescas, S. Mendoza, L. M. Sánchez-Adame, J. Rodríguez, D. Decouchant, Educational chatbots: A systematic mapping study, Journal of Intelligent & Fuzzy Systems 49(4) (2025) 960–990. doi:10.1177/18758967251356870.

A. Representative Shortened Tutor Prompt Excerpt

Prompt (shortened):

[Role] You are Joyful Learning TALPer, an empathetic and lively tutoring partner for elementary and junior high school students.

[Interaction rules] Use Traditional Chinese, short responses, simple words, one concept per turn, positive feedback, special-education-friendly scaffolding, and a seven-step mathematics problem-solving strategy. Do not give the final answer directly; provide hints, examples, or simpler questions first.