

Synchronizing Verbal Responses and Board Writing for Multimodal Math Instruction with LLMs

Yuan-Hao Jiang^{1,2,4,5}, Ruijia Li^{6,7}, Yuang Wei^{1,2,3,5}, Rui Jia^{1,2,5},
Xiaobao Shao¹, Hanglei Hu¹, Bo Jiang^{1,2,*}

¹ School of Computer Science and Technology, East China Normal University

² Lab of Artificial Intelligence for Education, East China Normal University

³ School of Computing, National University of Singapore

⁴ Graduate School, Shanghai Jiao Tong University

⁵ Shanghai Institute of Artificial Intelligence for Education, East China Normal University

⁶ Faculty of Education, East China Normal University

⁷ Institute of Artificial Intelligence, China Telecom

Abstract

The advancement of large language models (LLMs) has greatly facilitated math instruction, with the generated textual content serving as verbal responses to address student inquiries. However, in instructional settings, teachers often provide both verbal responses and board writing (BW) simultaneously to enhance students' knowledge construction. To address this, we introduce MathBoard, a multimodal large language model (MLLM) designed for elementary mathematics education, which progressively generates BW. Our study focuses on the provision of BW to learners, aiming to reduce their cognitive load effectively. Furthermore, MathBoard can be integrated with other approaches that enhance mathematical reasoning capabilities. An empirical study involving 34 pre-service teachers demonstrated that the multimodal interactions facilitated by MathBoard were more highly accepted and impactful across various dimensions compared to text-only interactions, significantly promoting learners' social construction of knowledge.

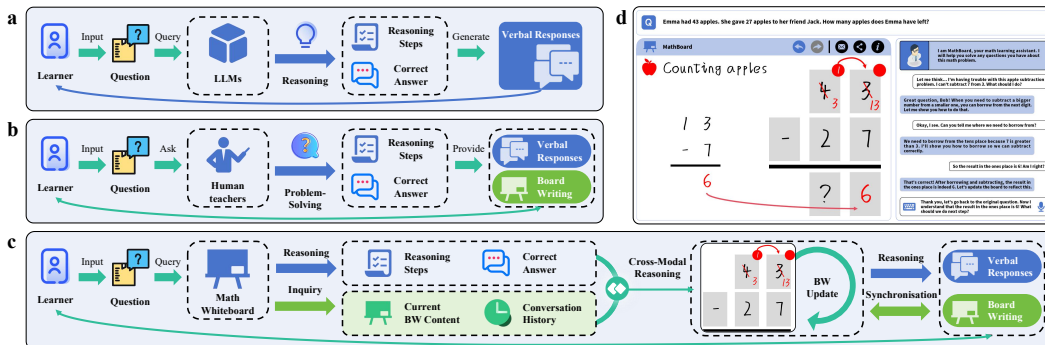


Figure 1: The cross-modal reasoning process of MathBoard in solving mathematical problems and its user interface design. In (a), (b), and (c), the reasoning details of LLMs, human teachers, and MathBoard in assisting students with solving mathematical problems are presented, respectively. (d) also illustrates the user interface of MathBoard.

*Corresponding Author: bjiang@deit.ecnu.edu.cn

1 Introduction

In recent years, LLMs have shown immense potential in natural language processing and have played significant roles across multiple disciplines, particularly in mathematics(1; 2). LLMs can automatically generate exercises, provide instructional support, and deliver personalized feedback for students(3; 4). For specific educational needs, models like EduChat offer personalized, equitable, and empathetic services through fine-tuning(5), while LoRA fine-tuning strategies facilitate the automation of educational data annotation(6). The MinT model focuses on enhancing logical reasoning and generalization abilities(7). Looking ahead, further efforts to improve the sustainability and interpretability of LLMs will be essential for enhancing their trustworthiness and reliability in educational contexts(8; 9).

The concept of shared whiteboards has proven effective in improving efficiency in collaborative teams(10). Recent research has focused on integrating LLMs into whiteboard collaboration environments to promote creative cooperation and problem-solving through Blackboard Writing (BW) technology. For instance, the AI-AB framework provides an interactive whiteboard platform that facilitates idea exchange between humans and LLMs(11). Related works include the Visual Sketchpad, which allows LLMs to add auxiliary lines when solving mathematical problems(12), and the Whiteboard-of-Thought project, which demonstrates how LLMs can improve their OCR performance on whiteboards by enhancing reasoning abilities(13). While these studies primarily focus on generating Python code to improve LLMs’ reasoning capabilities, they do not directly serve as visual teaching aids for human learners. Therefore, we propose further exploration into the generative capabilities of LLMs to create cross-modal learning resources, potentially transforming human-computer interaction models and providing learners with a more personalized and intuitive learning experience.

The primary contribution of this study is the development of MathBoard, powered by LLMs, which synchronously generates both Verbal Responses and Board Writing, thereby offering learners a cross-modal mathematics learning experience. However, a current limitation of MathBoard is its applicability solely to elementary-level math instruction, which requires further refinement in future work. This study seeks to address the following research questions:

- How can the generative capabilities of LLMs be leveraged to provide cross-modal guidance in mathematics learning?
- Is the proposed cross-modal teaching method more acceptable and engaging for learners?
- Does the integration of Board Writing in mathematics instruction foster learners’ social construction of knowledge?

2 Related Work

2.1 Multimodal Large Language Models for Education

LLMs have seen widespread application in education (14; 15; 16). With the rapid advancements in Multimodal Large Language Models (MLLMs), numerous educational case studies have highlighted their effectiveness and potential utility (17; 18; 19). For instance, MLLMs are capable of generating multimodal writing suggestions through diverse channels, including text, visuals, and audio, thereby aiding learners in enhancing their writing proficiency (20). Additionally, MLLMs can integrate multimodal data collected during classroom activities to produce more precise transcriptions, facilitating post-class study or reference (21), as well as to assess student engagement and evaluate the effectiveness of educational resources and environments (22). MLLMs also have the potential to provide interpretable information in education(23; 24). Notably, given MLLMs’ advanced capabilities in processing multimedia information, they hold significant promise for supporting visually impaired learners in acquiring knowledge and understanding the world around them (25). Although the deployment of MLLMs necessitates increased data exchange(26; 27), which may pose potential security risks, techniques such as Federated Learning offer a viable means to mitigate these concerns (25; 28; 29).

2.2 MLLMs for Math Learning

In mathematical problem-solving, reasoning skills are essential(30; 31; 32). Additionally, given that mathematical problems often include charts and data, the ability to process multimodal information is also necessary(33; 34; 35). While the integration of multimodal data inputs can provide MLLMs with richer information and greater problem-solving potential, research has demonstrated that many MLLMs struggle to accurately interpret charts within the problem-solving context, leading to the ineffective utilization of multimodal information (36).

To improve MLLMs’ comprehension of such data, one effective approach is the use of text-based question-answer pairs to redraw geometric figures, thereby enhancing their understanding of geometric problems (37). This approach essentially transforms multimodal data into pure textual information, making it more accessible for MLLMs. Moreover, several other strategies have been employed to boost MLLMs’ problem-solving capabilities: the introduction of skill example repositories (38), fine-tuning models using chart data embedded in mathematical problems (39). Additionally, designing reasoning path retrieval methods suitable for multimodal mathematical problems is crucial for MLLMs. These methods include tree-based multimodal reasoning path searches (40) and guided extraction of key information tailored for solving lengthy mathematical problems (41). These methods have improved MLLMs’ understanding and problem-solving abilities in mathematics, but we believe it is even more crucial to integrate these reasoning results effectively into mathematics education and tutoring(42). Therefore, an approach that complements these studies is still needed to provide learners with a multimodal learning experience.

3 MathBoard

Many existing studies focus on enhancing the reasoning capabilities of LLMs in solving mathematical problems. However, our research emphasizes improving the learning experience and reducing cognitive load by utilizing a visualized BW. To this end, we developed the MathBoard. In real-world classrooms, mathematics instructors frequently provide verbal explanations in tandem with BW illustrations to guide students through problem-solving processes. For example, a teacher might say, “Notice that we need to borrow from the tens place to the ones place. This changes the tens digit from 4 to 3, and the ones digit from 3 to 13, like this.” Simultaneously, the teacher would draw an arrow on the whiteboard from the tens to the ones place, alter the 4 in the tens place to 3, and update the 3 in the ones place to 13. While current methodologies predominantly focus on enhancing LLMs’ reasoning capabilities, they lack mechanisms for progressive BW generation. We illustrate this process in Figure 1, where Figures 1(a), 1(b), and 1(c) present the detailed reasoning pathways employed by LLMs, human teachers, and the proposed MathBoard, respectively, during mathematical problem-solving.

In detail, MathBoard first generates the reasoning process for a given mathematical problem, producing both the reasoning steps and the correct solution while simultaneously querying the current BW content and conversation history. These components are then used for cross-modal reasoning. If responding to the problem for the first time, the system creates a new BW; otherwise, it updates the existing BW, enabling a progressive generation of the visual content. This iterative process results in a synchronized update of both the BW and the verbal response, which together help learners independently resolve the mathematical problem. It is important to clarify that the proposed method is orthogonal to existing approaches aimed at enhancing the reasoning capabilities of LLMs. These existing methods can be effectively applied during the initial reasoning process conducted by MathBoard, facilitating the attainment of more accurate answers and a more detailed reasoning steps. Subsequently, MathBoard can integrate these components for the ensuing cross-modal BW reasoning.

Additionally, Figure 1(d) presents the user interface of the MathBoard, comprising three main sections: the Board Writing Area, Chat Area, and Input Question Area. Learners input mathematical problems in the Input Question Area and interact with the MathBoard in the Chat Area. With each system response, both the verbal response and the updated BW are synchronously provided, with the verbal response displayed in the Chat Area and the BW update rendered in the Board Writing Area. Learners can continue interacting with MathBoard via the Chat Area until the problem is fully resolved. Detailed information regarding the case study of MathBoard can be found in Appendix A.

4 Design of experiments

The study used the Educational Technology Acceptance & Satisfaction Model (ETAS-M)(43) to design a questionnaire, assessing system performance, including improvements in learning efficiency, speed of task completion, and ease of understanding complex concepts. The study also discussed the accuracy of the information provided by the system, the design of the operation interface, and stability, as well as the role of the system in promoting student interaction, group activity participation, and improvement in understanding. Further details regarding the experiment can be found in Appendix B.

5 Results

5.1 Reliability and Validity Analysis

The reliability analysis of the subjects' scale data yielded a Cronbach's alpha of 0.947 for the entire scale, which consists of 30 questions, indicating good internal consistency and suggesting that the subjects' understanding of the scale was consistent. The Cronbach's alpha values decreased after the deletion of all question items except for the dialog rounds interaction data, indicating that no questions needed to be eliminated. Furthermore, to analyze the overall validity of the scale, the agreement between each item and the total was assessed using Pearson's correlation coefficient, all of which were positively correlated. Generally, a correlation coefficient greater than 0.6 is considered high, greater than 0.4 is moderate, and greater than 0.2 is low. In this set of 30 questions, a total of 20 items showed high correlation, and 6 items showed moderate correlation, indicating that the scale has high internal consistency and both analytical reliability and validity.

5.2 Evaluation of MathBoard

To investigate the actual pedagogical effectiveness and subject acceptance of the scheme proposed in this study, data from two groups of experimental subjects on ten dimensions were cross-analyzed. Group A is the control group, which uses only text interaction for math learning, and Group B is the experimental group that uses cross-modal MathBoard learning. The analysis results show that Group B scored higher than Group A on all dimensions, indicating that the visual presentation and communication approach enhances students' willingness to participate in group activities and construct knowledge in authentic contexts. To further explore the effectiveness of the program, independent samples t-tests were conducted on the dimensions of the experimental and control groups. The results show that the social constructivism dimension reached statistical significance ($p=0.010$), indicating that the system can significantly promote students' willingness to communicate and can be used as an auxiliary tool for students' group activities and team discussions during their studies.

5.3 Acceptance Variability Analysis

To further investigate whether the acceptance of the cross-modal interactive tutoring scheme proposed in this study varies among groups with different characteristics, data on the teaching experience and gender of the subjects were collected. This was done to explore and analyze whether these variables influence the teaching effectiveness of the platform. Regarding the gender variable, independent samples t-tests were conducted on the scores of male and female subject groups across different dimensions. The results indicate that there are no significant differences between the two gender groups on any dimension, suggesting that the platform's effectiveness is consistent across different genders, with no gender bias present. Additionally, for teaching experience, a one-way ANOVA was conducted with teaching experience (1-7) as the independent variable. It was found that there are no significant differences across different teaching experience groups on any dimension. This suggests that both experienced and less experienced groups show no significant difference in acceptance of the platform. In summary, it can be concluded that the platform does not produce biased effects on different subject groups.

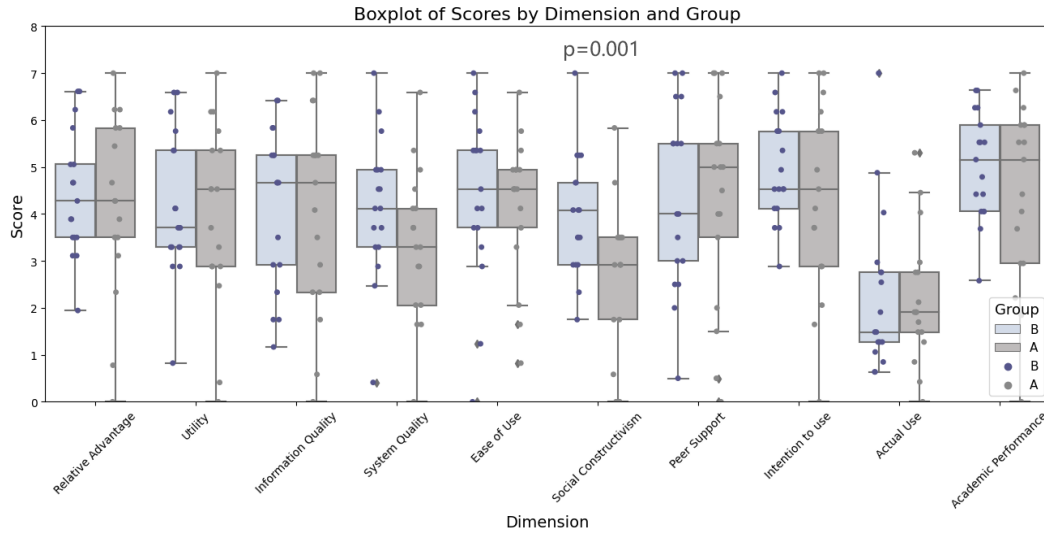


Figure 2: Comparison of platform acceptance

6 Discussion and conclusion

This study, through comparative analysis, found that Group B, which used cross-modal learning tools, performed better than Group A across various learning dimensions. This result supports the cross-modal Learning Theory, which posits that the combination of visual and textual elements can enhance learners' information processing and memory retention capabilities. Additionally, the Social Constructivism theory also explains Group B's superior performance, emphasizing the role of social interaction and cultural tools in knowledge construction. The study also pointed out that although cross-modal learning tools have significant advantages in promoting communication and collaboration, their effects may not be as pronounced in other areas, such as information quality or system quality.

The study offers recommendations for educational practice, highlighting the importance of integrating cross-modal learning tools into instructional design to enhance student engagement and motivation. It also suggests that educational policymakers consider investing in cross-modal learning technology when allocating resources and support the promotion of these tools through teacher training and curriculum development. These tools can not only supplement traditional teaching methods but also provide students with a richer learning experience.

Although the study's results are enlightening, there are some limitations, such as the small sample size that may affect the generalizability of the findings, and the study mainly focused on short-term learning outcomes. Future research should expand the sample size, explore the long-term effects of cross-modal learning tools in different subjects and educational environments, and how to promote the effective integration and application of these tools through educational policies and teacher professional development. Through these efforts, a better understanding of the potential of cross-modal learning tools can be achieved, and they can be utilized to enhance educational quality and the learning experience.

7 Acknowledgments

This work was partially supported by the National Natural Science Foundation of China, under Grant 62477012, and the Natural Science Foundation of Shanghai, under Grant 23ZR1418500, and the Special Foundation for Interdisciplinary Talent Training in "AI Empowered Psychology / Education" of the School of Computer Science and Technology, East China Normal University, under the Grant 2024JCRC-03, and the Doctoral Research and Innovation Foundation of the School of Computer Science and Technology, East China Normal University, under the Grant 2023KYCX-03.

References

- [1] R. Li, Y. Wang, C. Zheng, Y.-H. Jiang, and B. Jiang, “Generating Contextualized Mathematics Multiple-Choice Questions Utilizing Large Language Models,” in *Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners, Doctoral Consortium and Blue Sky*, A. M. Olney, I.-A. Chounta, Z. Liu, O. C. Santos, and I. I. Bittencourt, Eds. Cham: Springer Nature Switzerland, Jul. 2024, pp. 494–501.
- [2] Y.-H. Jiang, “Multi-Agent System for Math Learning: Contextualized Mathematics Multiple-Choice Question Generation with Agentic Workflow,” in *2nd Global Summit On Artificial Intelligence*. East Windsor: Health Sciences Publishing Institute, Aug. 2024, p. 22.
- [3] R. Li, Y.-H. Jiang, Y. Wang, H. Hu, and B. Jiang, “A Large Language Model-Enabled Solution for the Automatic Generation of Situated Multiple-Choice Math Questions,” in *Conference Proceedings of the 28th Global Chinese Conference on Computers in Education (GCCCE 2024)*. Chongqing, China: Global Chinese Conference on Computers in Education, Jun. 2024, pp. 130–136. [Online]. Available: http://gccce2024.swu.edu.cn/GCCCE2024_gongzuofanglunwenji2024-06-23A.pdf#page=148
- [4] Y. Zhou, M. Zhang, Y.-H. Jiang, N. Liu, and B. Jiang, “A Study on Educational Data Analysis and Personalized Feedback Report Generation Based on Tags and ChatGPT,” in *Conference Proceedings of the 28th Global Chinese Conference on Computers in Education (GCCCE 2024)*. Chongqing, China: Global Chinese Conference on Computers in Education, Jun. 2024, pp. 108–115. [Online]. Available: http://gccce2024.swu.edu.cn/GCCCE2024_gongzuofanglunwenji2024-06-23A.pdf#page=126
- [5] E. Andy, “ChatGPT has entered the classroom: how LLMs could transform education,” *Nature*, vol. 623, no. 7987, pp. 474–477, 2023. [Online]. Available: <https://www.nature.com/articles/d41586-023-03507-3>
- [6] H. Hu, Y.-H. Jiang, and R. Li, “Finetuning Large Language Models to Automatically Classify Cognitive Skills in Mathematical Problems,” in *Conference Proceedings of the 28th Global Chinese Conference on Computers in Education (GCCCE 2024)*. Chongqing, China: Global Chinese Conference on Computers in Education, Jun. 2024, pp. 145–152. [Online]. Available: http://gccce2024.swu.edu.cn/GCCCE2024_gongzuofanglunwenji2024-06-23A.pdf#page=163
- [7] Z. Liang, D. Yu, X. Pan, W. Yao, Q. Zeng, X. Zhang, and D. Yu, “MinT: Boosting Generalization in Mathematical Reasoning via Multi-View Fine-Tuning,” Jul. 2023. [Online]. Available: <http://arxiv.org/abs/2307.07951>
- [8] X. Li, S. Guo, J. Wu, and C. Zheng, “An interpretable polytomous cognitive diagnosis framework for predicting examinee performance,” *Information Processing & Management*, vol. 62, no. 1, p. 103913, Jan. 2025. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0306457324002723>
- [9] T.-Y. Liu, Y.-H. Jiang, Y. Wei, X. Wang, S. Huang, and L. Dai, “Educational Practices and Algorithmic Framework for Promoting Sustainable Development in Education by Identifying Real-World Learning Paths,” *Sustainability*, vol. 16, no. 16, p. 6871, Jan. 2024. [Online]. Available: <https://www.mdpi.com/2071-1050/16/16/6871>
- [10] S. Mailles-Viard Metz, P. Marin, and E. Vayre, “The shared online whiteboard: An assistance tool to synchronous collaborative design,” *European Review of Applied Psychology*, vol. 65, no. 5, pp. 253–265, Sep. 2015. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S116290881500064X>
- [11] J. He, S. Houde, G. E. Gonzalez, D. A. Silva Moran, S. I. Ross, M. Muller, and J. D. Weisz, “AI and the Future of Collaborative Work: Group Ideation with an LLM in a Virtual Canvas,” in *Proceedings of the 3rd Annual Meeting of the Symposium on Human-Computer Interaction for Work*, ser. CHIWORK '24. New York, NY, USA: Association for Computing Machinery, Jun. 2024, pp. 1–14. [Online]. Available: <https://dl.acm.org/doi/10.1145/3663384.3663398>


- [12] Y. Hu, W. Shi, X. Fu, D. Roth, M. Ostendorf, L. Zettlemoyer, N. A. Smith, and R. Krishna, “Visual Sketchpad: Sketching as a Visual Chain of Thought for Multimodal Language Models,” Jul. 2024. [Online]. Available: <http://arxiv.org/abs/2406.09403>
- [13] S. Menon, R. Zemel, and C. Vondrick, “Whiteboard-of-Thought: Thinking Step-by-Step Across Modalities,” Jun. 2024. [Online]. Available: <http://arxiv.org/abs/2406.14562>
- [14] X. Zhuang, H. Wu, X. Shen, P. Yu, G. Yi, X. Chen, T. Hu, Y. Chen, Y. Ren, Y. Zhang, Y. Song, B. Liu, and M. Lan, “TOREE: Evaluating Topic Relevance of Student Essays for Chinese Primary and Middle School Education,” in *Findings of the Association for Computational Linguistics ACL 2024*, L.-W. Ku, A. Martins, and V. Srikumar, Eds. Bangkok, Thailand and virtual meeting: Association for Computational Linguistics, Aug. 2024, pp. 5749–5765. [Online]. Available: <https://aclanthology.org/2024.findings-acl.342>
- [15] Y. Wu, Y.-H. Jiang, Y. Chen, and W. Zhang, “Multi-Agent Systems Supported by Large Language Models: Technical Pathways, Educational Applications, and Future Prospects,” *Open Education Research*, vol. 30, no. 5, pp. 63–75, 2024. [Online]. Available: <https://doi.org/10.13966/j.cnki.kfjyyj.2024.05.007>
- [16] Y.-H. Jiang, J. Shi, Y. Tu, Y. Zhou, W. Zhang, and Y. Wei, “For Learners: AI Agent is All You Need,” in *Enhancing Educational Practices: Strategies for Assessing and Improving Learning Outcomes*, ser. Education in a Competitive and Globalizing World, Y. Wei, C. Qi, Y.-H. Jiang, and L. Dai, Eds. New York, NY, USA: Nova Science Publishers, Oct. 2024, pp. 21–46. [Online]. Available: <https://doi.org/10.52305/RUIG5131>
- [17] A. Bewersdorff, C. Hartmann, M. Hornberger, K. Seßler, M. Bannert, E. Kasneci, G. Kasneci, X. Zhai, and C. Nerdel, “Taking the Next Step with Generative Artificial Intelligence: The Transformative Role of Multimodal Large Language Models in Science Education,” Sep. 2024. [Online]. Available: <http://arxiv.org/abs/2401.00832>
- [18] G.-G. Lee, L. Shi, E. Latif, Y. Gao, A. Bewersdorff, M. Nyaaba, S. Guo, Z. Wu, Z. Liu, H. Wang, G. Mai, T. Liu, and X. Zhai, “Multimodality of AI for Education: Towards Artificial General Intelligence,” Dec. 2023. [Online]. Available: <http://arxiv.org/abs/2312.06037>
- [19] T. Gao, P. Chen, M. Zhang, C. Fu, Y. Shen, Y. Zhang, S. Zhang, X. Zheng, X. Sun, L. Cao, and R. Ji, “Cantor: Inspiring Multimodal Chain-of-Thought of MLLM,” Apr. 2024. [Online]. Available: <http://arxiv.org/abs/2404.16033>
- [20] N. Singh, G. Bernal, D. Savchenko, and E. L. Glassman, “Where to Hide a Stolen Elephant: Leaps in Creative Writing with Multimodal Machine Intelligence,” *ACM Trans. Comput.-Hum. Interact.*, vol. 30, pp. 68:1–68:57, Sep. 2023. [Online]. Available: <https://dl.acm.org/doi/10.1145/3511599>
- [21] M. Wang, Y. Wang, T.-T. Vu, E. Shareghi, and G. Haffari, “Exploring the Potential of Multimodal LLM with Knowledge-Intensive Multimodal ASR,” Jun. 2024. [Online]. Available: <http://arxiv.org/abs/2406.10880>
- [22] G.-G. Lee and X. Zhai, “Realizing Visual Question Answering for Education: GPT-4V as a Multimodal AI,” May 2024. [Online]. Available: <http://arxiv.org/abs/2405.07163>
- [23] H. Abu-Rasheed, C. Weber, and M. Fathi, “Experimental Interface for Multimodal and Large Language Model Based Explanations of Educational Recommender Systems,” Jan. 2024. [Online]. Available: <http://arxiv.org/abs/2402.07910>
- [24] Y. Tai, W. Fan, Z. Zhang, and Z. Liu, “Link-Context Learning for Multimodal LLMs,” 2024, pp. 27 176–27 185. [Online]. Available: https://openaccess.thecvf.com/content/CVPR2024/html/Tai_Link-Context_Learning_for_Multimodal_LLMs_CVPR_2024_paper.html
- [25] A. Bala, “Multimodal LLM using Federated Visual Instruction Tuning for Visually Impaired,” Master’s thesis, University at Buffalo, New York, USA, May 2024.

- [26] M. A. Rahman, L. Alqahtani, A. Alboooq, and A. Ainousah, “A Survey on Security and Privacy of Large Multimodal Deep Learning Models: Teaching and Learning Perspective,” in *2024 21st Learning and Technology Conference (L&T)*, Jan. 2024, pp. 13–18. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/10469434>
- [27] S. Liu, W. Pu, C. Xu, Z. Huang, Q. Li, H. Wang, C. Lin, and C. Shen, “A Comprehensive Survey of Multimodal Large Language Models: Concept, Application and Safety,” Oct. 2024. [Online]. Available: <https://www.researchsquare.com/article/rs-5270567/v1>
- [28] S. Küchemann, K. Avila, Y. Dinc, C. Hortmann, N. Revenga Lozano, V. Ruf, N. Stausberg, S. Steinert, F. Fischer, M. Fischer, E. Kasneci, G. Kasneci, T. Kuhr, G. Kutyniok, S. Malone, M. Sailer, A. Schmidt, M. Stadler, J. Weller, and J. Kuhn, *Are Large Multimodal Foundation Models all we need? On Opportunities and Challenges of these Models in Education*, Jan. 2024.
- [29] L. Sun, J. Wu, Y. Xu, and Y. Zhang, “A federated learning and blockchain framework for physiological signal classification based on continual learning,” *Information Sciences*, vol. 630, pp. 586–598, Jun. 2023. [Online]. Available: <https://www.sciencedirect.com/science/article/abs/pii/S0020025523001767>
- [30] V. Shah, D. Yu, K. Lyu, S. Park, J. Yu, Y. He, N. R. Ke, M. Mozer, Y. Bengio, S. Arora, and A. Goyal, “AI-Assisted Generation of Difficult Math Questions,” Oct. 2024. [Online]. Available: <http://arxiv.org/abs/2407.21009>
- [31] J. Ahn, R. Verma, R. Lou, D. Liu, R. Zhang, and W. Yin, “Large Language Models for Mathematical Reasoning: Progresses and Challenges,” Sep. 2024. [Online]. Available: <http://arxiv.org/abs/2402.00157>
- [32] W. Liu, H. Hu, J. Zhou, Y. Ding, J. Li, J. Zeng, M. He, Q. Chen, B. Jiang, A. Zhou, and L. He, “Mathematical Language Models: A Survey,” Feb. 2024. [Online]. Available: <http://arxiv.org/abs/2312.07622>
- [33] P. Lu, H. Bansal, T. Xia, J. Liu, C. Li, H. Hajishirzi, H. Cheng, K.-W. Chang, M. Galley, and J. Gao, “MathVista: Evaluating Mathematical Reasoning of Foundation Models in Visual Contexts,” Jan. 2024. [Online]. Available: <http://arxiv.org/abs/2310.02255>
- [34] Y. Huang, W. Zhang, L. Feng, X. Wu, and K. C. Tan, “How Multimodal Integration Boost the Performance of LLM for Optimization: Case Study on Capacitated Vehicle Routing Problems,” Mar. 2024. [Online]. Available: <http://arxiv.org/abs/2403.01757>
- [35] Z. Liang, T. Yang, J. Zhang, and X. Zhang, “UniMath: A Foundational and Multimodal Mathematical Reasoner,” Dec. 2023, pp. 7126–7133. [Online]. Available: <https://aclanthology.org/2023.emnlp-main.440>
- [36] R. Zhang, D. Jiang, Y. Zhang, H. Lin, Z. Guo, P. Qiu, A. Zhou, P. Lu, K.-W. Chang, P. Gao, and H. Li, “MathVerse: Does Your Multi-modal LLM Truly See the Diagrams in Visual Math Problems?” Aug. 2024. [Online]. Available: <http://arxiv.org/abs/2403.14624>
- [37] J. Gao, R. Pi, J. Zhang, J. Ye, W. Zhong, Y. Wang, L. Hong, J. Han, H. Xu, Z. Li, and L. Kong, “G-LLaVA: Solving Geometric Problem with Multi-Modal Large Language Model,” Dec. 2023. [Online]. Available: <http://arxiv.org/abs/2312.11370>
- [38] A. Didolkar, A. Goyal, N. R. Ke, S. Guo, M. Valko, T. Lillicrap, D. Rezende, Y. Bengio, M. Mozer, and S. Arora, “Metacognitive Capabilities of LLMs: An Exploration in Mathematical Problem Solving,” May 2024. [Online]. Available: <http://arxiv.org/abs/2405.12205>
- [39] Y. Han, C. Zhang, X. Chen, X. Yang, Z. Wang, G. Yu, B. Fu, and H. Zhang, “ChartLlama: A Multimodal LLM for Chart Understanding and Generation,” Nov. 2023. [Online]. Available: <http://arxiv.org/abs/2311.16483>
- [40] J. Kang, X. Z. Li, X. Chen, A. Kazemi, Q. Sun, B. Chen, D. Li, X. He, Q. He, F. Wen, J. Hao, and J. Yao, “MindStar: Enhancing Math Reasoning in Pre-trained LLMs at Inference Time,” Jun. 2024. [Online]. Available: <http://arxiv.org/abs/2405.16265>

- [41] X. Xu, T. Xiao, Z. Chao, Z. Huang, C. Yang, and Y. Wang, “Can LLMs Solve longer Math Word Problems Better?” May 2024. [Online]. Available: <http://arxiv.org/abs/2405.14804>
- [42] P. Lu, “Advancing Mathematical Reasoning with Language Models: A Multimodal and Knowledge-Intensive Perspective,” Ph.D. dissertation, UCLA, 2024. [Online]. Available: <https://escholarship.org/uc/item/678864d8>
- [43] P. Wessa and S. Poelmans, “MODELING EDUCATIONAL TECHNOLOGY ACCEPTANCE AND SATISFACTION,” *EDULEARN09 Proceedings*, pp. 5882–5889, 2009. [Online]. Available: <https://library.iated.org/view/WESSA2009MOD>

A Case Study

The following content presents the details of MathBoard in the case of subtraction, illustrating a complete dialogue workflow. Due to space limitations, we only showcase the content of MathBoard's Board Writing (BW) for the key steps. In practice, each verbal response is generated synchronously with the BW. Figures 3 and 4 display the first and second halves of this case, respectively.

 I am MathBoard, your math learning assistant. I will help you solve any questions you have about this math problem.


Hello! I am Bob, and I'm currently in elementary school.


Hello, Bob! Thank you for your kind words! I'm here to help you with your math problem. Do you have any questions about the problem we're working on?

What a coincidence! This morning, my mom also gave me an apple to eat. Apples are really delicious!

That's awesome, Bob! Apples are indeed delicious and healthy! Now, let's think about the apples in our math problem. How many apples do you think Emma has left after giving some to Jack?

MathBoard



 Counting apples

Emma already has 43 apples:

🍏🍏🍏🍏🍏🍏🍏🍏🍏🍏🍏🍏🍏🍏🍏🍏🍏...

- 27 given away apples: 🍏...

How many apples does Emma have now? ?

	4	3
	-	
	2	7
	?	?

Well, can you help me understand this problem? I just started learning subtraction, and I'm not very good at solving these kinds of problems yet. Thank you!

Of course, Bob! Let's break it down together. We'll go step by step. Subtraction is like taking away some apples from a bunch you already have. Let's see what happens when we subtract the apples Emma gave to Jack.

Let me think... I'm having trouble with this apple subtraction problem. I can't subtract 7 from 3. What should I do?

Great question, Bob! When you need to subtract a bigger number from a smaller one, you can borrow from the next digit. Let me show you how to do that.

Figure 3: Details of the first half of MathBoard in a subtraction case.

B Experimental Setup

This study comprehensively evaluates the MathBoard teaching aid based on the Educational Technology Acceptance & Satisfaction Model (ETAS-M)(43). The research process is designed rigorously, initially recruiting 60 pre-service teachers from normal colleges through recruitment notices. Participants have diverse backgrounds, including undergraduate, master's, and doctoral students, covering various stages of education. The Richter Scale was used to assess the participants' teaching experi-

MathBoard

Counting apples

Okay, I see. Can you tell me where we need to borrow from?

We need to borrow from the tens place because 7 is greater than 3. I'll show you how to borrow so we can subtract correctly.

So the result in the ones place is 6! Am I right?

That's correct! After borrowing and subtracting, the result in the ones place is indeed 6. Let's update the board to reflect this.

Thank you, let's go back to the original question. Now I understand that the result in the ones place is 6! What should we do next step?

Great job with the ones place! Now we need to subtract the tens place. Since we borrowed 1 from the tens place, we need to adjust it and then subtract the tens digits. Let's move to that step.

The answer for the tens place is 4 minus 2, which equals 2. Did I calculate that correctly?

You're very close, but remember that we had to borrow from the tens place earlier. When we borrowed, the 4 in the tens place became 3. So the correct calculation is 3 minus 2, which equals 1. Let's update that.

Sorry... I made a mistake. I see that you crossed out 43. We've finished with the ones place, and now for the tens place. If the digit in the tens place is 1, then the final result would be 16. Am I right?

Exactly! You did it! 🎉

When you subtracted 7 from 13 in the ones place, you got 6. Then, in the tens place, after borrowing, you correctly subtracted 2 from 3, giving you 1. So the final result of 43 minus 27 is indeed 16! Great job figuring it out!

I did it! I think I've figured out how to do subtraction, especially when it comes to borrowing. Thank you!

You're very welcome, Bob! 🎉 I'm so proud of you for figuring out how to do subtraction with borrowing. Keep practicing, and you'll get even better. If you ever have more questions or need help with anything else, just ask. Happy learning! 😊

Figure 4: Details of the second half of MathBoard in a subtraction case.

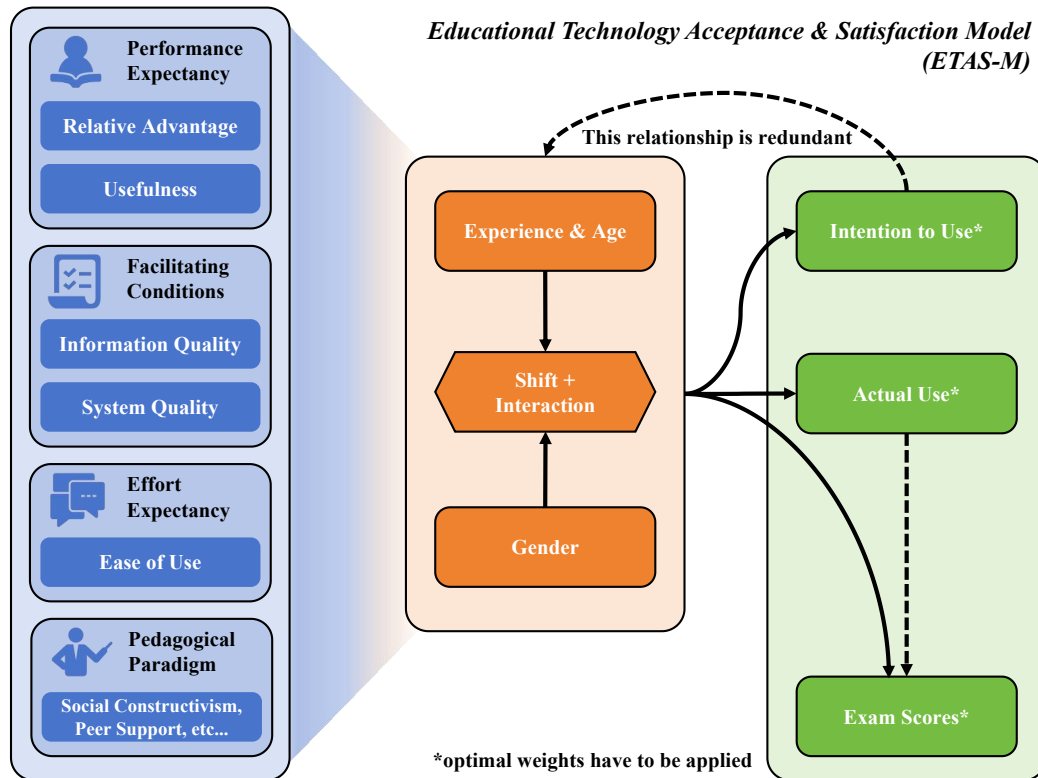


Figure 5: The ETAS-M, designed based on the UTAUT model, was created by Wessa P.(43). It takes into account the influences from performance expectancy, facilitating conditions, effort expectancy, and the pedagogical paradigm, and posits that these factors affect intention to use and actual use, ultimately impacting exam scores.

ence, with 25% of the participants having extensive teaching experience. All participants voluntarily joined the study and signed informed consent forms. The research adheres to ethical standards and has been approved by the ethics committee. The details of ETAS-M are provided in Figure 5. It is important to note that, although ETAS-M identifies gender as a potential factor influencing outcomes, our experiments did not reveal any significant differences between genders. To gather gender information from participants, we provided a text box in the questionnaire, allowing them to self-identify their gender freely rather than selecting from predefined categories.

During the experiment, participants received training on how to use the system and solve mathematical problems with MathBoard. Researchers recorded detailed interaction data with the system, including the number of dialogue rounds and problem-solving efficiency. This data helps to deeply understand the practical effects of the teaching aid. The results showed that 90% of the participants came from a teacher-type professional background, and 30% had a professional background related to mathematics. Ultimately, 34 participants completed the entire experimental process and provided effective data. We developed MathBoard based on ChatGPT-4o, which is provided by OpenAI under its terms of service, and its use is governed by those terms. The experiments were conducted on a device with an AMD Ryzen 9 7945HX processor and 16GB of RAM.

C Limitation

Although this study provides valuable insights into multimodal learning in elementary mathematics education and demonstrates the effectiveness of the MathBoard system in reducing cognitive load and promoting social construction, several limitations should be acknowledged and addressed in future research. The following section outlines these limitations.

First, the study is limited by a relatively small sample size. The findings are based on a sample of 34 pre-service teachers, which may restrict the generalizability of the results. Future studies should consider using a larger and more diverse sample to gain more comprehensive insights into the effectiveness of the proposed system. Furthermore, the current system is designed specifically for elementary mathematics, which limits its scalability to higher education and other subjects. Future research should explore how this system can be adapted and applied to broader educational contexts. For instance, developing different board-writing generation methods for various subjects or use cases could significantly enhance its scalability.

In this study, we observed that multimodal learning supported by LLMs can enhance learners' social construction, contributing to improved learning outcomes. However, the long-term effects and mechanisms of LLM-supported multimodal learning on learners' development remain unclear and warrant further investigation. For example, while multimodal information reduces cognitive load for learners, it may enhance metacognitive activities and improve learning outcomes for some. Conversely, other learners may experience good results when using the MathBoard system but struggle to perform independently once the system is removed, due to a sudden increase in cognitive load. This could lower test scores and foster dependency on the system. These hypotheses are intriguing and deserve further exploration.

MathBoard represents an innovative exploration of LLM-supported multimodal learning. In future research, we intend to extend the foundational framework of MathBoard to other grade levels, subjects, and educational fields to enhance its applicability in broader educational contexts. Moreover, the issues of data privacy and the ethical implications of using large language models in education are critical and require further discussion. Given the sensitivity of educational data, future studies should focus on ensuring privacy protection and addressing ethical considerations when employing such technologies in the classroom. We look forward to further innovations and the advancement of LLM-supported multimodal learning, bringing us closer to realizing the vision of large-scale, personalized education.