International Journal of Technology in Education
www.ijte.net

Artificial Intelligence in Mathematics Education: Assessing Accuracy, Reliability, and Instructional Implications

Eliseo P. Marpa

Philippine Normal University Visayas, Boy Albert St., Brgy. Zone 2, Cadiz City, Philippines, 0000-0001-9582-4333
Corresponding author: Eliseo P. Marpa (marpa.ep@pnu.edu.ph)

Article Info

Article History

Received:
27 July 2025

Revised:
3 January 2026

Accepted:
10 March 2026

Published:
18 June 2026

Keywords

Accuracy
Artificial intelligence
Mathematics learning
Reliability

Abstract

Artificial Intelligence (AI) is becoming increasingly embedded in educational practices, particularly in mathematics instruction. Its problem-solving capabilities and efficiency are explored even in a mathematics classroom. Thus, the study assesses pre-service mathematics teachers' perceptions of AI's accuracy, reliability, and implications for mathematics learning. A self-designed survey was administered to 128 randomly selected preservice mathematics teachers. The results revealed that AI Accuracy, in terms of solution correctness, problem complexity, speed and efficiency, error analysis, and domain-specific accuracy, was average. The same results were observed for reliability, including consistency, reproducibility, robustness, transparency, and error detection and correction. However, AI can be a valuable supplement to mathematics instruction. However, this should be handled with extra attention, not as it is. The findings suggest that AI may serve as a valuable supplement to teaching, but cannot yet substitute for the guidance and insight provided by human educators. This research highlights the importance of clear institutional policies and a critical evaluation of AI's role in supporting, rather than replacing, traditional instruction in mathematics learning.

Citation: Marpa, E. P. (2026). Artificial intelligence in mathematics education: Assessing accuracy, reliability, and instructional implications. *International Journal of Technology in Education (IJTE), 9*(3), 995-1006. https://doi.org/10.46328/ijte.5365

Introduction

Technological advancement continues to reshape educational environments, prompting educators and students to adapt to emerging tools such as Artificial Intelligence (AI). In education, AI is increasingly recognized for its potential to support instructional delivery, assessment, and personalized learning. Students frequently encounter AI-powered platforms inside and outside the classroom, and these systems are becoming integral to modern learning experiences (Holmes et al., 2022; Zawacki-Richter et al., 2019; Almuhanna, 2025).

Educational institutions and policy bodies worldwide, including those in the Philippines, are actively exploring the incorporation of AI into curricula. This shift underscores a growing need to understand the importance and limitations of AI applications in teaching and learning. According to recent initiatives by global education stakeholders (Cardona et al., 2023), AI integration is intended not only to enhance access to learning but also to innovate traditional instructional practices.

Existing literature suggests that AI can facilitate immediate feedback, adaptive support, personalized learning experiences, and collaborative learning opportunities for students when effectively integrated into instructional environments (Lin et al., 2023; Ouyang & Zhang, 2024). Intelligent tutoring systems, chatbots, and virtual assistants exemplify how these technologies might improve engagement and foster deeper cognitive processing. Nevertheless, concerns remain regarding the extent to which AI tools can support complex reasoning and problem-solving, particularly in a discipline as intricate as mathematics.

Empirical studies have highlighted both the potential and limitations of artificial intelligence in education. Research indicates that AI-driven adaptive learning systems can enhance student achievement and engagement by tailoring instruction to individual learning needs and providing timely feedback (Wang et al., 2024; Lin et al., 2023). At the same time, scholars emphasize that the effectiveness of AI depends largely on how these technologies are integrated into pedagogical practices and supported by human oversight (Holmes et al., 2022; Seo et al., 2025).

The findings reflect a broader trend that while AI can enhance certain aspects of instruction, it cannot yet replicate the nuanced decision-making, contextual understanding, and responsiveness of human teachers (Holmes et al., 2022; Zawacki-Richter et al., 2019). In mathematics education, the demand for systems that promote conceptual understanding, abstract reasoning, and critical thinking remains high. AI-powered tools can provide adaptive problem sets, intelligent tutoring, interactive simulations, and personalized feedback that support mathematical exploration and learning (Awang et al., 2025; Lin et al., 2023).

However, current systems still struggle with deeper interpretive tasks and non-routine problem-solving. As noted by Wu (2021) and Voskoglou and Salem (2020), while AI tools may aid comprehension, their effectiveness hinges on how they are implemented and contextualized within broader teaching strategies. In light of this, the present research seeks to assess the accuracy, reliability, and implications for mathematics learning of AI as perceived by pre-service mathematics teachers. This work also considers how such technologies might influence the future of mathematics education, particularly regarding their role as instructional supports rather than replacements for human educators.

Theoretical Bases

This study is anchored in three interrelated theoretical perspectives: constructivist learning theory, cognitive load theory, and the emerging framework of human-AI partnership in education. Constructivist theory posits that learners construct their understanding through active engagement with content and social interaction (Piaget, 1972; Vygotsky, 1978). In the context of AI-assisted mathematics learning, constructivism supports the use of tools that encourage inquiry, exploration, and reflective thinking rather than passive information intake. AI systems that provide interactive simulations or problem-solving feedback align with these principles by helping learners build conceptual knowledge through engagement with dynamic content.

Cognitive load theory (Sweller, 1988) offers additional insight into how technology can support or hinder learning. According to this framework, instructional design should minimize unnecessary cognitive effort, allowing learners to focus on relevant problem-solving tasks. Adaptive AI systems can reduce extraneous cognitive load by tailoring content difficulty, offering immediate clarification, and automating repetitive tasks. However, if poorly implemented, such systems may introduce distractions or overly simplify tasks, limiting deep learning.

Complementing these theories is the human-AI collaboration framework, which emphasizes the need for a balanced interplay between computational intelligence and human judgment (Seo et al., 2025; Meylani, 2024). In education, this approach highlights how AI can augment teacher roles without replacing them (Seo et al., 2025; Kolchenko, 2018). AI tools are most effective when they assist teachers in assessment, provide personalized learning support, or automate routine tasks, allowing educators to concentrate on higher-level instructional responsibilities (Kokku et al., 2018; Van der Vorst & Jelicic, 2019; Maghsudi et al., 2021). Together, these frameworks guide the interpretation of AI's effectiveness in mathematics education, not merely in terms of speed or automation, but in how it contributes to meaningful learning experiences (Maghsudi et al., 2021; Holmes et al., 2022). The evaluation of AI’s accuracy and reliability in this study is thus grounded in pedagogical considerations that prioritize learner autonomy, instructional quality, and cognitive engagement (Holmes et al., 2022; Seo et al., 2025).

Method

Research Design

This study utilized a descriptive quantitative research design to evaluate the perceived accuracy, reliability, and implications of Artificial Intelligence (AI) tools in mathematics education. Descriptive research is appropriate for studies aiming to capture and interpret current perceptions or conditions without manipulating variables or establishing causal relationships (Fraenkel & Wallen, 2012; Gall & Borg, 2007). Given the study's focus on how pre-service mathematics teachers assess AI-based systems, this design enabled a detailed examination of participants' insights from structured survey data. The descriptive approach enabled the researcher to explore the current state of AI integration in educational practice without intervening in instructional processes.

Sample Size, Sample, and Sampling Technique

The study involved 128 pre-service mathematics education students enrolled at the Philippine Normal University, Visayas. The participants were selected through random sampling**,** ensuring that every eligible student had an equal chance of being included. This sample size was determined using G*Power 3.1, a statistical power analysis tool widely used in the social sciences. The power analysis was conducted to ensure adequate statistical reliability and minimize the risk of Type II errors (Faul, Erdfelder, Buchner, & Lang, 2009). G*Power enabled the researcher to compute the minimum required sample size based on expected effect sizes, power level (0.80), and significance threshold ($\alpha = 0.05$), supporting the credibility of the study's findings.

Research Instruments

Data were collected through a *researcher-developed survey instrument* composed of three major sections:

Part I captured participants' demographic and academic profiles.

Part II assessed perceptions of AI *accuracy* in mathematics learning. This part included 25 items categorized into five domains: correctness of solutions, complexity of problems, speed and efficiency, error analysis, and domain-specific accuracy.

Part III measured the perceived reliability of AI systems using an additional 25 items across five domains: consistency, reproducibility, robustness, transparency, and error detection and correction.

Each item used a five-point Likert scale, ranging from 1 (strongly disagree) to 5 (strongly agree), allowing respondents to indicate the extent to which they agreed with each statement.

Validity and Reliability of the Instruments

To establish content validity, the survey instrument was reviewed by five subject-matter experts in mathematics education, research methodology, and educational technology. Reviewers examined each item to determine whether the item is essential, essential but not important, or not essential. Based on reviewers' feedback, several items were retained, and others were revised. The developed research instrument Content Validity Ratio (CVR) was 0.99, indicating a high level of validity and appropriateness for use in the study. The research instrument's reliability was evaluated through a pilot test with 32 students enrolled in a similar program who were excluded from the main study. Using Cronbach's alpha, the overall reliability coefficients were calculated as 0.93 for the accuracy domain and 0.95 for the reliability domain. These values exceed the commonly accepted threshold of 0.70 for educational research, indicating a very high degree of reliability (Gliem & Gliem, 2003).

Statistical Data Analysis

The data collected were analyzed using descriptive statistical methods**,** specifically the mean and standard deviation, to evaluate participants' perceptions of AI systems in the context of mathematics education. These measures were employed to assess central tendencies and variability in responses related to the accuracy and reliability of AI tools when applied to tasks such as problem-solving, feedback generation, and performance consistency

Results and Discussion

This section presents the survey findings, structured around three central themes: the accuracy and reliability of AI, and AI's implications for mathematics learning. Each theme is discussed in relation to the research objectives, with attention to both confirming and contradictory responses across participant groups.

The Accuracy of AI when Applied to Tasks Related to Learning Mathematics

Table 1 summarizes respondents' perceptions of AI's accuracy when applied to tasks related to mathematics learning. The overall mean (M = 3.05, SD = 0.49) indicates that AIs are perceived as performing at an average level of accuracy. This finding suggests that while AI systems can be useful in routine mathematical applications, they do not consistently excel across all functional areas. The relatively low standard deviation reflects general agreement among participants. However, this average performance also points to the existing limitations of AI in addressing deeper conceptual tasks and in supporting nuanced forms of mathematical reasoning.

Table 1. The Level of Accuracy of AI when Applied to Mathematics Learning

Accuracy	Mean	Sd	Interpretation
Correctness of Solutions	2.93	0.64	Moderate
Complexity of Problems	2.77	0.61	Moderate
Speed and Efficiency	3.34	0.67	Moderate
Error Analysis	3.09	0.55	Moderate
Domain-Specific Accuracy	3.07	0.58	Moderate
As a Whole	**3.05**	**0.49**	**Moderate**

When assessing solution correctness, the results indicate a moderate mean score (M = 2.94, SD = 0.63). This implies that while AI tools can often deliver correct answers, their reliability is not absolute. Learners may encounter variability, especially with open-ended problems or questions requiring multiple solution steps. According to Davis (2023), AI still struggles to handle word problems that require integrating mathematical operations with real-world contexts. Gorban et al. (2018) have proposed error-filtering mechanisms using linear discriminants, particularly effective in high-dimensional data environments. Such tools may enhance accuracy, but current implementations in education remain imperfect.

In terms of problem complexity, AI systems were rated lower (M = 2.78, SD = 0.60), reflecting student perceptions that these tools are less capable of solving tasks that involve abstract thinking or multiple cognitive steps. As problems become more intricate, students find that AI responses may lose coherence or relevance. Funke et al.

(2018) emphasized that solving complex problems requires both cognitive modeling and strategic reasoning, attributes that current AI systems do not fully replicate. The lower mean in this category points to a gap between computational output and conceptual understanding.

AI systems performed best in speed and efficiency, which received the highest average score (M = 3.33, SD = 0.68). Respondents recognized AI's strength in rapidly processing mathematical inputs and producing quick responses. However, despite this speed, accuracy, and explanation quality remain moderate. Studies have highlighted that AI possesses significant computational advantages in processing large volumes of information and automating routine tasks; however, efficiency and speed alone do not guarantee meaningful learning outcomes, which depend on instructional design, learner engagement, and pedagogical quality (Holmes et al., 2022; Maghsudi et al., 2021).

The relatively high standard deviation indicates that students' experiences with AI’s efficiency are not uniform, likely influenced by the specific platform or task type. The error analysis results were also in the moderate range (M = 3.08, SD = 0.56). This suggests that AI tools are somewhat capable of identifying mistakes, but may not always offer clear or pedagogically sound corrections. While many systems can highlight errors, their feedback is often limited to surface-level issues rather than deeper misconceptions. Ozkaya (2020) notes that many machine learning-based AI tools are inherently probabilistic, making their behavior less predictable and harder to verify. Recent research suggests that advances in meta-learning and adaptive AI systems can improve model performance, generalization, and responsiveness to new tasks, offering promising directions for future educational AI applications (Hospedales et al., 2021).

Lastly, domain-specific accuracy received a similar moderate evaluation (M = 3.08, SD = 0.59). This reflects a perception that AI performs better within narrowly defined mathematical topics but less effectively across integrated or interdisciplinary domains. Hendrycks et al. (2021), using the MATH dataset, demonstrated that even large AI models fail to consistently solve high-school-level math problems, suggesting that mere scale does not guarantee reasoning ability. Domain-specific strengths mirror patterns observed in human learning, where expertise in one area does not always generalize across others (Scott & Berman, 2013). Researchers such as Islam et al. (2022) and Hwang & Tu (2021) continue to explore AI systems tailored to learners’ specific needs through personalization and explainable logic.

The Reliability of AI when Applied to Tasks Related to Learning Mathematics

The study assessed how pre-service mathematics students perceived the reliability of AI systems in performing math-related instructional tasks. As shown in Table 2, the overall reliability was evaluated as moderate (M = 3.02, SD = 0.58). This average rating reflects a cautious level of trust in AI's consistency and functionality. The relatively low standard deviation suggests minimal variability in student perceptions, indicating that most respondents shared similar experiences. However, this midpoint score also reveals that AI systems are not yet widely viewed as highly dependable across diverse educational contexts. Factors such as task complexity, AI design variability, and user interaction may contribute to these limitations.

Table 2. The Reliability of AI when Applied to Mathematics Learning

Reliability	Mean	Sd	Interpretation
Consistency	3.08	0.67	Moderate
Reproducibility	2.96	0.77	Moderate
Robustness	2.97	0.65	Moderate
Transparency	3.05	0.58	Moderate
Error detection and correction	3.01	0.72	Moderate
As a Whole	**3.02**	**0.58**	**Moderate**

In terms of consistency, AI systems received the highest score among the five indicators (M = 3.08, SD = 0.67). This suggests that, on average, students found AI capable of producing stable results under similar conditions. While some fluctuations exist, as reflected in the moderate spread, learners generally perceived AI as predictable when faced with familiar mathematical problems. However, this does not imply full confidence. Prior research by Davies et al. (2021) indicates that even consistent outputs do not always translate into reliable reasoning, especially in tasks requiring abstraction or multi-step solutions.

The dimension of reproducibility was rated lower (M = 2.96, SD = 0.77), suggesting uncertainty about whether AI can yield similar outcomes across different scenarios or users. The relatively high standard deviation here suggests more varied responses, likely due to inconsistent experiences with AI tools. This mirrors broader concerns raised by Gundersen and Kjensmo (2018), who found that reproducibility in AI research is compromised by undocumented processes and variations in model training. Although AI shows promise in content delivery and automation, this result reinforces the fact that it still lacks uniformity in its application.

When evaluating robustness, participants also gave an average rating (M = 2.97, SD = 0.65). Students viewed AI systems as moderately capable of handling new or unfamiliar inputs, but not reliably so. The relatively narrow standard deviation suggests that most participants shared this view. This finding supports the work of Corso et al. (2023), who emphasized that while some models demonstrate robustness to adversarial changes or dataset shifts, most educational AI tools struggle to generalize beyond their programmed scope. For transparency, AI systems also received a moderate score (M = 3.04, SD = 0.57). Respondents generally agreed that AI systems offer a certain level of traceability in how they generate solutions, but not to a fully satisfactory degree. The standard deviation shows that opinions were mostly aligned. This supports the insights of Schmidt et al. (2020), who noted that although transparency is seen as essential for trust, it is not always perceived as helpful when paired with inconsistent outputs. In practice, students may find it difficult to understand the logic behind AI decisions, limiting their educational value.

The ability of AI systems to detect and correct errors was also assessed as average (M = 3.01, SD = 0.72). This indicates that students view AI as somewhat capable of identifying mistakes, though with varying effectiveness. The relatively wide spread in responses suggests inconsistencies in perceived reliability, possibly due to differences in task complexity or the type of AI tool used. According to Blood et al. (2023) and Kim et al. (2021), enhancing AI's diagnostic functions requires both technical refinement and user-centered design. Current systems may detect surface-level errors but lack the deeper analytical capacity to provide corrective feedback aligned with mathematical reasoning.

Implications for Mathematics Learning

This study set out to examine how artificial intelligence (AI) systems are perceived in terms of accuracy and reliability when applied to mathematics learning. The results contribute to a growing body of research suggesting that while AI can enhance certain aspects of instruction, particularly in processing speed and operational efficiency, it remains limited in its ability to support higher-order mathematical reasoning. These limitations emphasize the continued importance of human teachers in guiding conceptual understanding and critical thinking in mathematics.

The findings point to several opportunities for future research and development. One area of focus should be improving AI's ability to navigate more complex and abstract mathematical problems, which are often less structured and demand interpretive judgment. Additionally, strengthening AI's capacity to provide accurate, pedagogically meaningful feedback, especially in detecting and correcting errors, will be essential to maximizing its instructional value. Further studies may also explore how AI tools can be designed to function more effectively alongside human-led instruction, supporting rather than replacing teacher expertise.

From a policy and institutional perspective, these findings underscore the need for well-defined strategies around the use of AI in classrooms. As adoption increases, educational stakeholders must ensure that AI applications are pedagogically aligned, technically reliable, and implemented in ways that complement teacher instruction. Policymakers should invest in systems tailored specifically to the challenges of mathematics education, particularly those designed to enhance deep learning, conceptual exploration, and student engagement with non-routine problems. This study does present some limitations. The focus on pre-service mathematics teachers may not reflect the full diversity of AI users in educational contexts. Moreover, because the data were based on user perceptions rather than direct technical assessments, the results reflect experiential perspectives rather than empirical system evaluations. To build on these findings, future research could incorporate objective performance metrics or real-time classroom implementations to measure the actual impact of AI tools.

Generally, while AI technologies show promise in enhancing mathematics instruction, especially in task automation and basic procedural support, they are not yet sufficient to replace the instructional roles that require interpretation, scaffolding, and adaptive feedback. Teachers remain central to the learning process, and as AI capabilities continue to evolve, their integration must be approached with care, ensuring that technology strengthens rather than substitutes the core functions of human instruction.

Conclusions and Recommendations

The findings show that artificial intelligence systems in mathematics education are perceived to perform at a moderate level of accuracy and reliability across key dimensions, including accuracy, problem complexity, processing speed, and error identification. While participants acknowledged improvements in efficiency and responsiveness, they also expressed concerns about the system's capacity to manage complex, abstract, or open-ended mathematical tasks that require reasoning beyond procedural execution. These outcomes highlight the complementary role AI tools may play in classroom instruction. Specifically, AI appears well-suited to assist with routine computations, provide real-time feedback, and, to some extent, personalize learning pathways. However, its limitations in supporting higher-order thinking, conceptual clarity, and adaptive instruction confirm that human educators remain vital to the teaching and learning process.

The study adds to ongoing scholarly discourse regarding both the promise and constraints of AI in educational settings. While AI can increase access to learning tools and streamline repetitive instructional tasks, it has yet to demonstrate consistent reliability or depth in addressing the cognitive demands of mathematics. Therefore, strategic improvements in AI design, particularly in problem-solving ability, diagnostic feedback, and contextual learning support, are necessary for broader, more meaningful classroom integration. Future research should prioritize developing AI that is more context-aware and pedagogically aligned. Investigations could also explore hybrid models in which AI works in tandem with educators to reinforce learning objectives and scaffold student understanding. Moreover, longitudinal studies examining AI's impact on actual learning outcomes across diverse student populations would help deepen the evidence base.

While AI continues to evolve and demonstrates tangible benefits for certain aspects of mathematics education, it does not yet possess the interpretive or instructional depth to replace teachers. Human facilitation remains critical, particularly for tasks that require conceptual understanding, error diagnosis, and adaptive teaching strategies. Ongoing collaboration between technologists, educators, and researchers will be essential to fully realize the potential of AI as a supportive, not substitutive, educational tool.

Statements and Declarations

Acknowledgments/Notes: The author extends deep gratitude to all who contributed to the completion of this research. Sincere appreciation goes to the preservice teachers specializing in mathematics, whose participation made this study possible. Heartfelt thanks are also due to the author's colleagues for their assistance in administering the research instruments. The support and encouragement from family members have been invaluable throughout this journey. The author is equally thankful to the Philippine Normal University, Visayas, for providing the platform and resources needed to pursue this work. Above all, profound thanks are offered to God Almighty for strength, wisdom, and grace, without which this endeavor would not have been possible.

During the preparation of this article, the authors did not use ChatGPT.

Supplementary Materials: Not applicable.

Author Contributions: The author has read and agreed to the published version of the manuscript.

Funding: The author received no funding for the research.

Data Availability: Not applicable.

Ethics Approval: The study was performed in accordance with the study protocol and ethical guidelines and regulations.

Informed Consent: Informed consent was obtained from all subjects involved in the study.

Conflicts of Interest: The author declares no conflicts of interest.

References

Almuhanna, M. A. (2025). Teachers' perspectives of integrating AI-powered technologies in K–12 education for creating customized learning materials and resources. *Education and Information Technologies.* https://doi.org/10.1007/s10639-024-13257-y

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*(1). https://doi.org/10.29333/iejme/16006

Blood, J.C., Herbert, N., & Wayne, M.R. (2023). Reliability Assurance for AI Systems. *2023 Annual Reliability and Maintainability Symposium (RAMS)*, 1-6.

Cardona, M. A., Rodríguez, R. J., & Ishmael, K. (2023). *Artificial intelligence and the future of teaching and learning: Insights and recommendations.* U.S. Department of Education, Office of Educational Technology.

Corso, A., Karamadian, D., Valentin, R., Cooper, M., & Kochenderfer, M.J. (2023). A Holistic Assessment of the Reliability of Machine Learning Systems. *ArXiv, abs/2307.10586*.

Davis, E. (2023). Mathematics, word problems, common sense, and artificial intelligence. *ArXiv, abs/2301.09723*.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Fraenkel, J. R., & Wallen, N. E. (2012). *How to design and evaluate research in education* (8th ed.). McGraw-Hill.

Funke, J., Fischer, A., & Holt, D. V. (2018). Competencies for complexity: Problem solving in the twenty-first century. In E. Care, P. Griffin, & M. Wilson (Eds.), *Assessment and teaching of 21st century skills: Research and applications* (pp. 41–53). Springer. https://doi.org/10.1007/978-3-319-65368-6_3

Gall, M. D., Gall, J. P., & Borg, W. R. (2007). *Educational research: An introduction* (8th ed.). Pearson.

Gliem, J. A., & Gliem, R. R. (2003). Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for Likert-type scales. *Midwest Research-to-Practice Conference in Adult, Continuing, and Community Education.*

Gorban, A. N., Golubkov, A., Grechuk, B., Mirkes, E. M., & Tyukin, I. (2018). Correction of AI systems by linear discriminants: Probabilistic foundations. *Information Sciences, 466*, 303-322. DOI: 10.1016/j.ins.2018.07.040

Gundersen, O., & Kjensmo, S. (2018). State of the Art: Reproducibility in Artificial Intelligence. *AAAI Conference on Artificial Intelligence*.

Hendrycks, D., Burns, C., Kadavath, S., Arora, A., Basart, S., Tang, E., Song, D.X., & Steinhardt, J. (2021). Measuring Mathematical Problem Solving with the MATH Dataset. *ArXiv, abs/2103.03874*.

Holmes, W., Bialik, M., & Fadel, C. (2022). *Artificial intelligence in education: Promises and implications for teaching and learning.* Center for Curriculum Redesign.

Hospedales, T., Antoniou, A., Micaelli, P., & Storkey, A. (2021). Meta-learning in neural networks: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 44*(9), 5149–5169. https://doi.org/10.1109/TPAMI.2021.3079209

Hwang, G., & Tu, Y. (2021). Roles and Research Trends of Artificial Intelligence in Mathematics Education: A Bibliometric Mapping Analysis and Systematic Review. *Mathematics*.

Islam, M.R., Ahmed, M.U., Barua, S., & Begum, S. (2022). A Systematic Review of Explainable Artificial Intelligence in Terms of Different Application Domains and Tasks. *Applied Sciences*.

Kim, K. J., & Han, H. J. (2021). *A design and effect of maker education using educational artificial intelligence tools in elementary online environment.* Journal of Digital Convergence.

Kolchenko, V. (2018). Can modern AI replace teachers? Not so fast! Artificial intelligence and adaptive learning: Personalized education in the AI age. *HAPS Educator, 22*(3), 249–252.

Kokku, R., Sundararajan, S., Dey, P., Sindhgatta, R., Sengupta, B., & Chakraborty, B. (2018). Augmenting classrooms with AI for personalized education. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 6976–6980). IEEE.

Lin, C. C., Huang, A. Y. Q., & Lu, O. H. T. (2023). Artificial intelligence in intelligent tutoring systems toward sustainable education: A systematic review. *Smart Learning Environments, 10*(41). https://doi.org/10.1186/s40561-023-00260-y

Maghsudi, S., Lan, A., Xu, J., & van der Schaar, M. (2021). Personalized education in the artificial intelligence era: What to expect next. *IEEE Signal Processing Magazine, 38*(3), 37–50. https://doi.org/10.1109/MSP.2021.3055702

Maghsudi, S., Lan, A., Xu, J., & van der Schaar, M. (2021). Personalized education in the artificial intelligence era: What to expect next. *IEEE Signal Processing Magazine, 38*(3), 37–50. https://doi.org/10.1109/MSP.2021.3055702

McMaster, M.D. (2008). Resolving Complex Problems.

Meylani, R. (2024). Artificial intelligence in the education of teachers: A qualitative synthesis of the cutting-edge research literature. *Journal of Computer and Education Research.*

Ouyang, F., & Zhang, L. (2024). AI-driven learning analytics applications and tools in computer-supported collaborative learning: A systematic review. *Educational Research Review, 43*, 100613. https://doi.org/10.1016/j.edurev.2024.100613

Ozkaya, I. (2020). What Is Really Different in Engineering AI-Enabled Systems? *IEEE Softw., 37*, 3-6.

Piaget, J. (1972). *The psychology of the child.* Basic Books.

Schmidt, P., Biessmann, F., & Teubner, T. (2020). Transparency and trust in artificial intelligence systems. *Journal of Decision Systems, 29*, 260 - 278.

Scott, B.M., & Berman, A. (2013). Examining the Domain-Specificity of Metacognition Using Academic Domains and Task-Specific Individual Differences. *Australian Journal of Educational and Developmental Psychology, 13*, 28-43.

Seo, K., Yoo, M., Dodson, S., & Jin, S. H. (2025). Augmented teachers: K–12 teachers' needs for artificial intelligence's complementary role in personalized learning. *Journal of Research on Technology in Education.*

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Van der Vorst, T., & Jelicic, N. (2019). *Artificial intelligence in education: Can AI bring the full potential of personalized learning to education?* European Parliament Policy Department for Economic, Scientific and Quality of Life Policies.

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

Wang, X., Huang, R. T., Sommer, M., Pei, B., & others. (2024). The efficacy of artificial intelligence-enabled adaptive learning systems from 2010 to 2022 on learner outcomes: A meta-analysis. *Journal of Educational Computing Research.* https://doi.org/10.1177/07356331241240459

Wu, R. (2021). Visualization of basic mathematics teaching based on artificial intelligence. Journal of Physics: Conference Series, 1992(1), 042042. https://doi.org/10.1088/1742-6596/1992/4/042042

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—Where are the educators? *International Journal of Educational Technology in Higher Education, 16*(39). https://doi.org/10.1186/s41239-019-0171-0