Kalamatika: Jurnal Pendidikan Matematika

P-ISSN 2527-5615
E-ISSN 2527-5607

Journal homepage: https://kalamatika.matematika-uhamka.com/

Exploring the impact of artificial intelligence on mathematics learning: A meta-analysis

Harun Onesimus Laia, Bambang Avip Priatna Martadiputra*, and Endang Cahya Mulyaning Asih, Yosy Candraningsih

To cite this entry:

Laila, H. O., Martadiputra, B. A. P., Asih, E. C. M., & Candraningsih, Y. (2026). Exploring the impact of artificial intelligence on mathematics learning: A meta-analysis. *Kalamatika: Jurnal Pendidikan Matematika, 11*(2), 197-215. https://doi.org/10.22236/KALAMATIKA.vol11no2.2026pp197-215

Link to the article online:
https://kalamatika.matematika-uhamka.com/index.php/kmk/article/view/815

Submitted: *Nov 25, 2025* | Revised*: Mar 17, 2026* | Accepted: *Mar 19, 2026* |
Published online: *November 22, 2026*

© 2026 The Author(s). Published by FKIP Universitas Muhammadiyah Prof. Dr. HAMKA in collaboration with Indonesian Mathematics Educators (IMES).

This work is an Open Access article licensed under a Creative Commons Attribution - ShareAlike 4.0 International License.

OPEN ACCESS

Kalamatika: Jurnal Pendidikan Matematika
Volume 11, No. 2, November 2026, pages 197-215
DOI: 10.22236/KALAMATIKA.vol11no2.2026pp197-215

Exploring the impact of artificial intelligence on mathematics learning: A meta-analysis

Harun Onesimus Laia, Bambang Avip Priatna Martadiputra, Endang Cahya Mulyaning Asih, and Yosy Candraningsih

Department of Mathematics Education, Universitas Pendidikan Indonesia, Bandung, Indonesia

ABSTRACT

Despite the growing integration of artificial intelligence (AI) in education, empirical findings regarding its effectiveness in mathematics learning remain mixed, providing limited insight into the conditions under which AI-based interventions are most effective. This study aims to examine the effectiveness of AI in mathematics learning and investigate whether educational level, intervention duration, type of AI, mathematical competence, and mathematical content area moderate its effectiveness. A meta-analysis was conducted using 37 effect sizes from 16 studies that met the inclusion criteria. The studies were identified through a systematic search of the Scopus database, and the data were analysed using a random-effects model and Hedges' g, with the support of Comprehensive Meta-Analysis (CMA) software. The findings indicate that AI implementation has a statistically significant positive effect on mathematics learning ($g = 0.65$, $p < 0.05$). The moderator analysis revealed that mathematical competence and educational level were significant moderators, whereas intervention duration, type of AI, and mathematical content area were not. These findings highlight the importance of aligning AI-based interventions with the targeted mathematical competencies and learners' educational characteristics. This study contributes to the literature by synthesizing empirical evidence on the effectiveness of AI in mathematics learning and identifying key moderating factors that may influence its impact.

KEYWORDS

Artificial intelligence; mathematics learning; meta-analysis

CORRESPONDING AUTHOR

Bambang Avip Priatna Martadiputra ✉ bambangavip@upi.edu 🖃 Department of Mathematics Education, Universitas Pendidikan Indonesia, Bandung, Indonesia

INTRODUCTION

Artificial intelligence (AI) has increasingly emerged as a pivotal force in reshaping various aspects of human life, including education. Its rapid development has enabled the creation of learning environments that are more adaptive, personalised, and driven by continuous data processing. In mathematics education, AI has been implemented to provide automated feedback, deliver guidance tailored to students' levels of understanding, and strengthen conceptual learning through tools such as Intelligent Tutoring Systems (ITS), Adaptive Learning Systems (ALS), natural language processing (NLP)-based chatbots, and educational robots (Al-Nawaiseh et al., 2024; Berida et al., 2024; Ellington, 2026; Lu et al., 2021). These technologies facilitate the interactive and dynamic delivery of mathematical content, which can be adapted to accommodate learners' individual needs. As educational practices continue to shift towards digital formats, AI is increasingly regarded as an innovative approach to addressing various challenges in mathematics instruction (Baker et al., 2019). Its

use is expected to enhance students' conceptual understanding, problem-solving skills, and engagement in mathematics learning (Awang et al., 2025; Memiş, 2025; Yıldız & Körpeoğlu, 2025).

Various types of AI have demonstrated positive potential in mathematics learning, particularly in supporting cognitive development and enhancing students' mathematical competencies, which may contribute to improved learning outcomes. For instance, Intelligent Tutoring Systems (ITS) support students by providing personalised learning pathways, adaptive instructional guidance, and real-time feedback based on their progress and performance (Lu et al., 2021). Adaptive Learning Systems (ALS) promote student engagement by incorporating game-based learning features and offering learning pathways that dynamically adjust to individual learners' progress (Akavova et al., 2023). Meanwhile, robotics-based AI provides interactive, hands-on learning experiences that can foster academic achievement, critical thinking, and collaboration (Lampropoulos & Papadakis, 2024). These findings suggest that AI is not merely a learning tool but can also serve as an effective instructional facilitator.

Nevertheless, research examining the effectiveness of AI in mathematics education has yielded varied and sometimes conflicting results. Some studies have reported significant improvements in students' mathematics achievement following AI-based interventions (Khazanchi, 2024; Hankel, 2024; Alvarez, 2024), whereas others have found only minimal or statistically insignificant effects (Hwang, 2022; Yi et al., 2025). This variation may be influenced by several contextual factors that could moderate the effectiveness of AI, including intervention duration, the type of AI technology used, the mathematical content taught, the type of mathematical competence targeted, and students' educational levels. For example, the context and subject matter of learning strongly influence how technology is applied and may lead to different learning outcomes. Consequently, the impact of AI may vary depending on the mathematical topic being taught (D'Mello & Graesser, 2015; Hwang, 2022). Similarly, differences in intervention duration may result in varying effects on learning outcomes (Létourneau et al., 2025; Wu, 2024).

The types of AI used in mathematics education also vary, and these differences may influence learning outcomes. Intelligent Tutoring Systems (ITS) focus on monitoring students' progress and adjusting learning pathways according to their individual needs (Lu et al., 2021). Adaptive Learning Systems (ALS) modify instructional content in real time to accommodate individual learning needs (Chen et al., 2020). Meanwhile, AI-powered robots serve as interactive agents that can promote student engagement (Belpaeme et al., 2018). Each type of

AI employs different mechanisms and instructional approaches, which may lead to varying effects on students' mathematical competencies. Furthermore, the specific type of mathematical competence being targeted may also moderate the effectiveness of AI. Competencies such as conceptual understanding, problem-solving, and critical thinking may respond differently to AI-based interventions because educational technologies can influence cognitive processes and learning outcomes in different ways, depending on the competencies being developed (Garcia et al., 2026). Educational level may also influence the effectiveness of AI in mathematics learning because learners' cognitive characteristics and learning experiences vary across stages of schooling. These differences may shape how students interact with and benefit from technology-enhanced learning environments (Niño et al., 2023; Zhang et al., 2025). Therefore, educational level should be considered when evaluating the impact of AI on mathematics learning outcomes.

Given these considerations, a comprehensive synthesis of the effects of AI on mathematics learning is essential, particularly to identify the factors that may moderate its effectiveness. These moderating factors have not been extensively examined in previous meta-analytic studies focusing specifically on mathematics education. Previous studies have generally adopted a narrower scope, frequently concentrating on a single educational level, such as elementary education (Hwang, 2022), or on specific types of AI or mathematical topics (Fardian et al., 2025). This study seeks to systematically evaluate the effectiveness of AI in mathematics learning by examining five potential moderating factors: intervention duration, type of AI, mathematical competence, mathematical content area, and students' educational level. These factors were selected because differences in implementation duration, technological characteristics, learning objectives, subject matter, and learners' developmental stages may shape how AI is integrated into instruction and how students interact with AI-supported learning environments. Considering these dimensions allows the analysis to capture the complexity of AI-based instruction, as its impact may depend on the duration of the intervention, the characteristics of the technology used, the competencies learners are expected to develop, and the educational context in which learning takes place (Garcia et al., 2026; Létourneau et al., 2025; Wu, 2024). By examining these factors, the present study aims to provide a more comprehensive and nuanced synthesis of empirical evidence, thereby contributing to a deeper understanding of the conditions under which AI produces stronger or weaker effects on mathematics learning. This broader perspective may help explain inconsistencies in previous findings and provide a stronger evidence base for educators,

technology developers, and policymakers seeking to design AI-supported learning environments that are targeted, adaptive, and responsive to students' needs across different instructional settings.

METHODS

This study employed a meta-analytic approach (Borenstein et al., 2021) to systematically examine the effect of artificial intelligence (AI) on students' mathematics learning outcomes. In addition, several potential moderators were coded and analysed, including educational level, intervention duration, type of AI, mathematical competence, and mathematical content area, to identify factors that may influence the effectiveness of AI in mathematics learning. According to Borenstein et al (2021), meta-analysis is a quantitative technique that combines findings from multiple empirical studies to draw more comprehensive conclusions. An essential step in meta-analytic research is conducting a thorough and systematic literature search to identify studies relevant to the research objectives. Following the initial identification of potentially relevant studies, a screening process was conducted using predefined inclusion and exclusion criteria. These criteria provided a standardised framework for evaluating the relevance and suitability of each study and determining its eligibility for inclusion in the final meta-analytic dataset.

The inclusion and exclusion criteria were established to ensure that the selected studies were relevant and suitable for meta-analysis. The inclusion criteria were as follows: (1) publications released between 2015 and 2025; (2) articles published in academic journals or conference proceedings; (3) articles containing statistical information that allowed effect sizes to be calculated or extracted for meta-analysis; (4) articles examining the application of AI in mathematics education; and (5) studies for which the full text was accessible. Studies were excluded if they were literature reviews, including systematic reviews or meta-analyses; theoretical articles without empirical data; studies that lacked sufficient statistical information for effect-size calculation; or duplicate publications. Applying these criteria helped ensure that the data synthesis was based on relevant and methodologically suitable studies.

Data were collected through an online search for previous studies on AI in mathematics education using the Scopus database, which was selected because of its broad coverage of peer-reviewed academic publications. Focusing on a single database also helped maintain consistency in the study identification and selection process. The search terms were adapted to the research topic and included relevant synonyms and related terms. The search string was as follows:

("artificial intelligence" OR "AI in education" OR "intelligent tutoring system" OR "machine

learning" OR "chatbot") AND (mathematics OR "mathematics achievement" OR "math performance" OR "mathematical ability" OR "mathematics learning outcomes").

The retrieved articles were screened to ensure that only studies meeting all the inclusion criteria were selected for further analysis, thereby enhancing the relevance of the final dataset (Onesimus Laia et al., 2025). The data were analysed using Hedges' *g*, a standardised effect-size measure commonly used in meta-analysis to account for potential bias associated with differences in sample sizes across studies (Khairunnisa et al., 2022). Effect sizes based on Hedges' *g* (Thalheimer & Cook, 2002) were classified into five categories: ES < 0.15 (negligible), 0.15 ≤ ES < 0.40 (low), 0.40 ≤ ES < 0.75 (moderate), 0.75 ≤ ES < 1.10 (high), and ES ≥ 1.10 (very high). Publication bias was assessed using funnel plots and the Duval and Tweedie Trim and Fill procedure to identify and adjust for potential publication bias (Duval & Tweedie, 2000). Heterogeneity was assessed using the Q statistic and its corresponding *p*-value to evaluate variability in the effect sizes across studies. When substantial heterogeneity was identified, moderator analyses were conducted to examine factors that might account for differences in the effectiveness of AI. All analyses were performed using Comprehensive Meta-Analysis (CMA) software.

RESULTS AND DISCUSSION

The study selection process began with an initial search that identified 451 studies. Title and abstract screening resulted in the exclusion of 356 studies, leaving 95 studies for further evaluation. A subsequent full-text assessment based on the inclusion and exclusion criteria resulted in the exclusion of a further 79 studies: 18 were irrelevant to the research topic, 27 employed unsuitable methodologies, 7 lacked sufficient statistical data, and 27 were inaccessible. Ultimately, 16 studies met all the inclusion criteria, yielding 37 effect sizes related to the implementation of AI in mathematics learning for the meta-analysis. Figure 1 presents the PRISMA flow diagram illustrating the study selection process.

Kalamatika: Jurnal Pendidikan Matematika
Volume 11, No. 2, November 2026, pages 197-215
DOI: 10.22236/KALAMATIKA.vol11no2.2026pp197-215

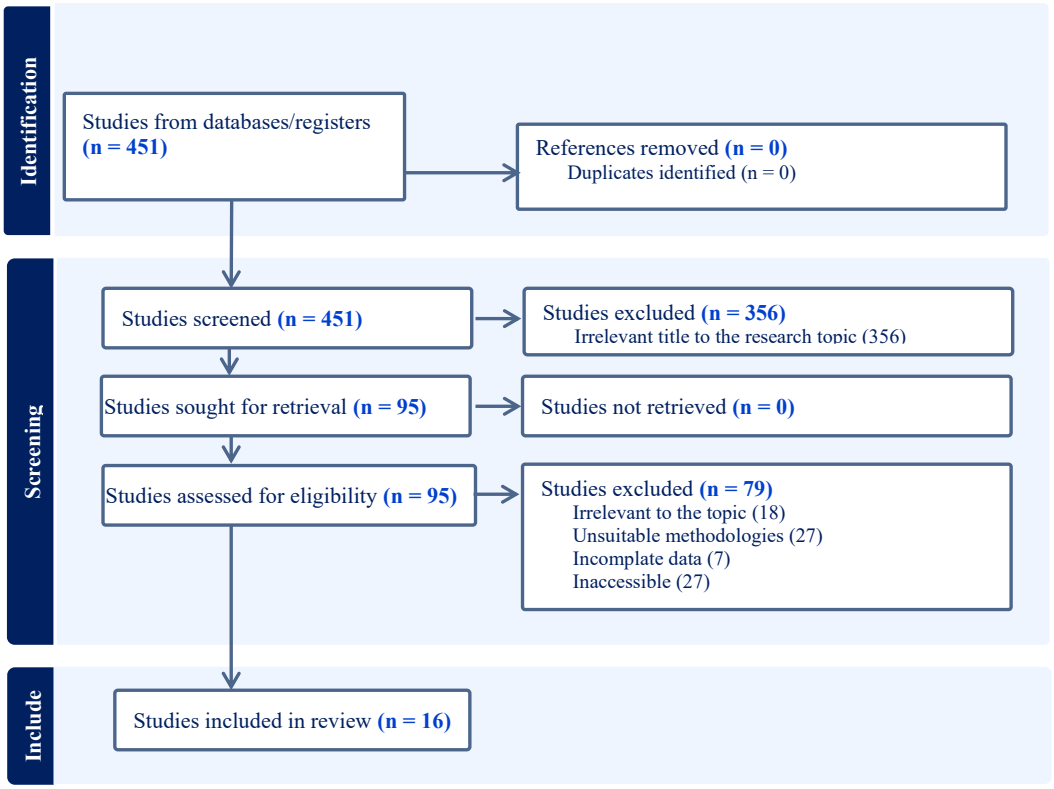


Figure 1. PRISMA Flow Diagram

To provide a clearer overview of the studies selected through the process illustrated in Figure 1, the characteristics of the 16 included articles are summarised in Table 1.

Table 1. Included Studies

Author(s)	Title	Year
Candace Walkington & Matthew L. Bernacki	Personalising algebra to students' individual interests in an intelligent tutoring system: moderators of impact	2018
Conrad Borchers et al.	What makes problem-solving practice effective? Comparing paper and AI tutoring	2023
Avnee Gaur & Karuna Kalita	Impact of enhanced learning approaches on stem-focused education for school children in Assam, India	2025
Fadi Bani Ahmad et al.	The effect of interactive ai tools like Photomath on developing mathematical concepts in students with learning difficulties: A quasi-experimental study	2025
Fatima Wahba et al.	The impact of ChatGPT-based learning statistics on undergraduates' statistical reasoning and attitudes toward statistics	2024
Ibrahim A. H. El-Shara et al.	The effect of using MathGPT on mathematical proficiency among undergraduate students	2025
Javier del Olmo-Muñoz et al.	Using intra-task flexibility on an intelligent tutoring system to promote arithmetic problem-solving proficiency	2022
Jenny Nehring et al.	Assessing the effectiveness of an artificial intelligence tutoring system for improving college-level mathematics preparedness in high school students	2023
Joel I. Alvarez	Evaluating the impact of AI–powered tutors MathGPT and Flexi 2.0 in enhancing calculus learning	2024
Manolis Mavrikis et al.	Short-term efficacy and usage recommendations for a large-scale	2018

Author(s)	Title	Year
	implementation of the Math-Whizz tutor	
Mohammed A. Alomari & Manar Omar Jabr	The effect of the use of an educational software based on the strategy of artificial intelligence on students' achievement and their attitudes towards it	2020
Nejla Gürefe et al.	Investigating the impact of the AI-supported 5E (AI-S5E) instructional model on spatial ability	2024
Niko Rummel et al.	Combining exploratory learning with structured practice to foster conceptual and procedural fractions knowledge	2016
Rashmi Khazanchi et al.	The effect of AI-based systems on mathematics achievement in rural context: A quantitative study	2025
Seth King et al.	An automated virtual reality training system for teacher-student interaction: A randomized controlled trial	2022
Shu-Chuan Shih et al.	Mathematics intelligent tutoring system for learning multiplication and division of fractions based on diagnostic teaching	2023

Following the initial screening and selection of studies, publication bias was assessed to evaluate the robustness of the meta-analytic findings. This assessment is important for determining whether the results may be disproportionately influenced by the selective publication of studies based on the characteristics or statistical significance of their findings. As shown in Figure 2, a funnel plot was used to visually examine the distribution of effect sizes across the included studies. The symmetry and spread of the funnel plot were then evaluated to identify potential asymmetry, which may indicate the presence of publication bias.

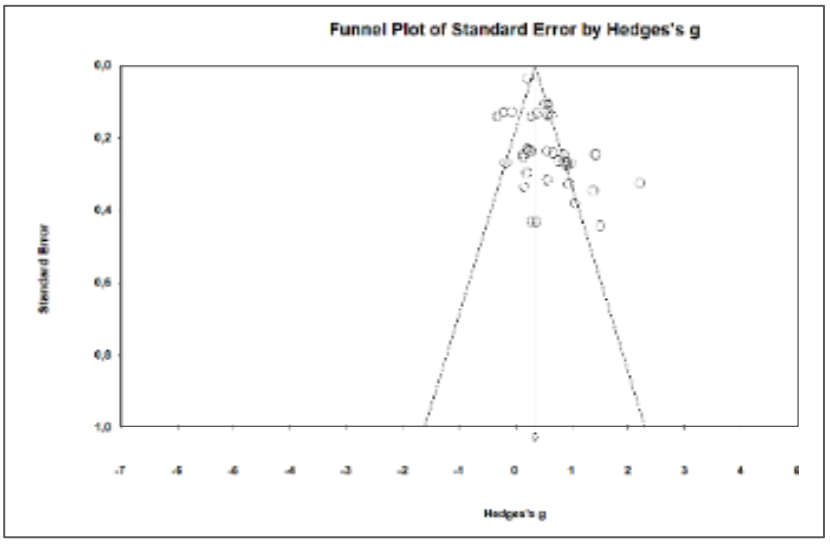


Figure 2. Funnel Plot

Figure 2 presents a funnel plot in which the data points appear relatively symmetrical around the mean effect size, suggesting a relatively low likelihood of publication bias (Sterne et al., 2011). To further investigate the potential influence of publication bias, the Duval and Tweedie Trim and Fill procedure was applied. This method estimates the potential impact of

Kalamatika: Jurnal Pendidikan Matematika
Volume 11, No. 2, November 2026, pages 197-215
DOI: 10.22236/KALAMATIKA.vol11no2.2026pp197-215

studies that may be missing from the dataset due to publication bias. This additional analysis provides further insight into whether the observed effect sizes may have been inflated or distorted by potentially missing studies.

Table 2. The Results of Fill and Trim Test

Model		Fixed Effects			Random Effects			Heterogeneity		
	ST	PE	LL	UL	PE	LL	UL	*Q*-value	*Df* (*Q*)	*p*-value
Observed values	0	0.339	0.291	0.387	0.649	0.477	0.822	310.1	36	0,000
Adjusted values		0.339	0.291	0.387	0.649	0.477	0.822			

ST: Studies Trimmed
PE: Point Estimated
LL: Lower Limit
UL: Upper Limit

According to Table 2, the Duval and Tweedie Trim and Fill analysis yielded a "Studies Trimmed" value of 0, indicating that no potentially missing studies were estimated by the procedure and that the pooled effect-size estimate remained unchanged (Duval & Tweedie, 2000). The heterogeneity analysis yielded a Q statistic of 310.1, indicating substantial variation in effect sizes across the included studies and supporting the use of a random-effects model (Sedgwick, 2015). This substantial heterogeneity warrants further investigation through subgroup or moderator analyses (Ruppar, 2020). Therefore, examining potential moderating factors is appropriate.

Figure 3 presents a forest plot illustrating the individual and overall effect sizes, together with their corresponding confidence intervals. The distribution of effect sizes shows that 7 were classified as negligible, 9 as low, 7 as moderate, 8 as high, and 6 as very high. The pooled effect size estimated using the random-effects model was 0.65, indicating a moderate effect. The statistical significance test ($p < .001$) demonstrates that AI use has a statistically significant positive effect on mathematics learning. This finding is consistent with Vygotsky's Zone of Proximal Development, which emphasises the importance of adaptive scaffolding, such as that potentially provided by AI systems (Choi & McClenen, 2020). AI can facilitate individualised learning experiences, provide immediate feedback, and enhance student engagement, all of which may contribute to improved learning outcomes (Awang et al., 2025; Opesemowo, 2025). The observed effect can also be discussed in relation to cognitive load theory, which suggests that although AI may support learner autonomy and improve learning efficiency, learning outcomes remain influenced by students' cognitive capacity to process complex information independently (Sweller, 2011).

Kalamatika: Jurnal Pendidikan Matematika
Volume 11, No. 2, November 2026, pages 197-215
DOI: 10.22236/KALAMATIKA.vol11no2.2026pp197-215

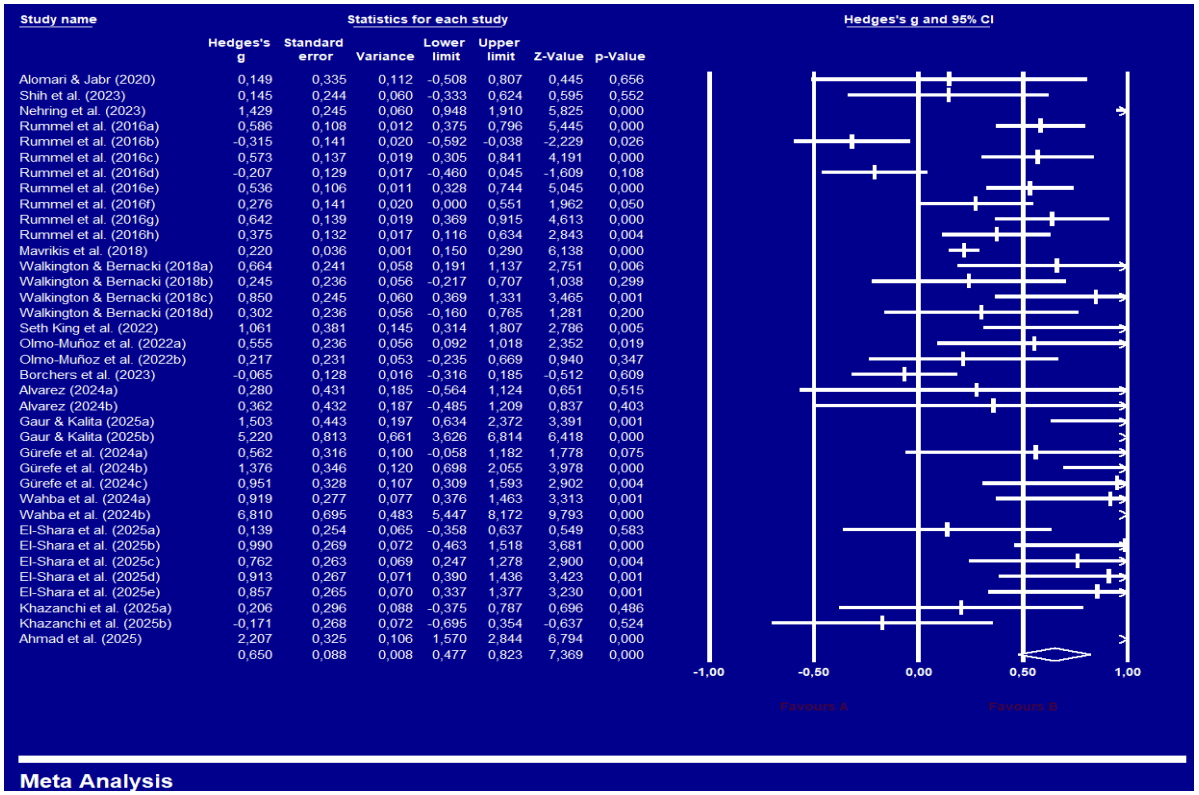


Figure 3. Forest Plot

Given the heterogeneity across studies, Table 3 presents the moderator analysis conducted to better understand the factors that may influence the observed effects.

Table 3. Moderator analysis

Moderator Variable	Subgroup	Effect size		Test of null (2-Tail)		Heterogeneity		
		Number of studies	Point estimate	Z-value	P-value	Q-value	*Df*	*p*-value
Educational Level	Elementary	10	0.434	3.505	0.000			
	Middle	11	0.517	3.013	0.003			
	High	5	0.695	3.250	0.001	9.547	4	0.049
	College	10	1.150	3.719	0.001			
	Mixed	1	1.153	3.391	0.001			
Intervention Duration	0-5 Weeks	22	0.690	5.136	0.000			
	6-10 Weeks	6	0.565	3.334	0.001	0.444	2	0.801
	Over 10 Weeks	9	0.733	3.372	0.001			
AI Type	ALS	7	0.745	2.364	0.018			
	ITS	28	0.543	6.193	0.000	2.553	2	0.279
	Robotics	2	3.299	1.776	0.076			
Mathematical Competence	Achievement	7	1.112	2.936	0.003			
	Computational	2	3.299	1.776	0.076			
	Conceptual	6	0.447	1.759	0.079			
	Problem Solving	11	0.371	3.532	0.000	14.194	6	0.028
	Procedural	5	0.509	5.601	0.000			
	Reasoning	2	0.916	4.764	0.000			
	Spatial	3	0.947	4.050	0.000			
Mathematical Content Area	Algebra	8	0.427	2.239	0.025			
	Calculus	7	0.660	4.939	0.000			
	Geometry	3	0.947	4.050	0.000	4.929	4	0.295
	Numbers	17	0.524	4.866	0.000			
	Statistics	2	3.830	1.301	0.193			

Educational Level

The educational-level moderator was categorised into five groups. The results showed that the estimated effectiveness of AI in mathematics learning increased with educational level. The largest effect sizes were observed in the college (1.150) and mixed (1.153) categories, followed by high school (0.695), middle school (0.517), and elementary school (0.434). The Q statistic of 9.547, with a *p*-value of 0.049, indicates that educational level is a statistically significant moderator. This finding suggests that learners at higher educational levels may have more developed self-regulated learning (SRL) skills, including metacognitive awareness, planning, and self-evaluation, which may enable them to manage their learning more independently, including when using AI-based learning tools (Seban & Urban, 2024; Zimmerman, 2002). In addition, Sweller (2011) emphasised that effective learning requires appropriate cognitive load management. Students at higher educational levels may therefore possess more complex prior knowledge schemas, allowing AI-based tools to provide instructional support that is better aligned with their cognitive needs. Conversely, elementary school students may require simpler, more visual, and interactive approaches when using AI-based learning tools because they are still developing their abstract reasoning abilities and tend to benefit from visual representations and guided support when learning mathematical concepts. Overall, the effectiveness of AI in mathematics learning may be influenced by learners' cognitive preparedness and digital competence across different educational levels.

Intervention Duration

The moderator analysis for intervention duration indicated that this variable was not a statistically significant moderator of the effect of AI on mathematics learning ($Q = 0.444$, $df = 2$, $p = 0.801$). Although the estimated effect sizes varied slightly across intervention durations (0–5 weeks: 0.690; 6–10 weeks: 0.565; more than 10 weeks: 0.733), these differences were not statistically significant. This suggests that, within the range examined in this study, intervention duration does not substantially moderate the overall effect of AI on mathematics learning outcomes. This finding may be explained by the importance of instructional design quality and the nature of interactions with technology, rather than the duration of exposure alone. For instance, principles of effective instructional design suggest that well-structured learning experiences that manage cognitive load and promote active engagement can produce favourable learning outcomes across varying intervention durations (Surbakti et al., 2024). Furthermore, Al-Nawaiseh et al. (2024) emphasised that instructional design quality, implementation context, and students' adaptability to technology are key factors in the effective integration of AI into

educational settings, including mathematics education.

AI Type

The meta-analysis showed that all three types of AI had positive estimated effects on mathematics learning. Robotics produced the largest estimated effect size (3.299), followed by Adaptive Learning Systems (ALS) (0.745) and Intelligent Tutoring Systems (ITS) (0.543). Although the result for robotics indicates a large positive estimated effect, the limited number of studies in this category means that this finding should be interpreted cautiously. Further primary research is needed to provide stronger empirical evidence. The Q statistic of 2.553 and the *p*-value of 0.279 indicate that the differences among AI types were not statistically significant. Thus, AI type was not identified as a statistically significant moderator of effectiveness. Nevertheless, the variation in estimated effect sizes may reflect differences in the pedagogical approaches associated with each AI type. Robotics-based AI tends to offer more interactive and concrete learning experiences, potentially enhancing motivation and engagement, particularly in STEM contexts (Ouyang & Xu, 2024). In contrast, ITS and ALS focus more on personalisation and automated feedback, although their effectiveness may depend heavily on the quality of the instructional content and the adaptive algorithms employed (Holmes et al., 2019). Therefore, although AI type was not a statistically significant moderator, it remains an important consideration in instructional planning because different AI types offer distinct instructional approaches that may support the development of specific mathematical skills.

Mathematical Competence

The moderator analysis indicated that mathematical competence was a statistically significant moderator of the effect of AI on mathematics learning ($Q = 14.194$, $p = 0.028$). This finding suggests that the estimated effects of AI varied according to the type of mathematical competence targeted. Positive estimated effects were observed for Achievement (1.112), Problem Solving (0.371), Procedural (0.509), Reasoning (0.916), and Spatial (0.947) competencies. The largest effect sizes among these categories were observed for Spatial and Reasoning competencies. This may be attributable to AI's capacity to provide dynamic visualisations, opportunities for object manipulation, and adaptive scaffolding, which may facilitate the development of higher-order cognitive abilities (Gürefe et al., 2024; Wahba et al., 2024). Although Computational competence had a very large estimated effect size (3.299), this result was not statistically significant ($p = 0.076$). A similar pattern was observed for

Conceptual competence (0.447, $p = 0.079$). These non-significant findings may be related to the small number of studies available for these categories, particularly Computational competence ($N = 2$). The relatively smaller effect sizes for Problem Solving and Conceptual competencies may suggest that developing these competencies requires more complex interactions or divergent thinking processes that are not yet fully supported by current AI technologies.

Mathematical Content Area

The moderator analysis for mathematical content area indicated that this variable was not a statistically significant moderator of the effect of AI on mathematics learning ($Q = 4.929$, $p = 0.295$). This finding suggests that the effectiveness of AI did not differ significantly across the mathematical content areas examined. Nevertheless, most content areas demonstrated positive estimated effects: Algebra (0.427), Calculus (0.660), Geometry (0.947), and Numbers (0.524). The Statistics category showed a relatively large estimated effect size (3.830), although this result was not statistically significant ($p = 0.193$). This finding was based on a very limited number of studies ($N = 2$), indicating that the available evidence remains preliminary and that further primary research is needed to provide stronger empirical support. Overall, these findings suggest that core AI capabilities, such as adaptive practice, personalised feedback, and conceptual visualisation, may be applied across various mathematical domains. This is consistent with learner-centred instructional principles, according to which AI can adapt to individual learning needs without being restricted to specific content areas (Baker & Siemens, 2014).

Despite these findings, this study has several limitations. First, the limited number of primary studies included in the meta-analysis may restrict the generalisability of the findings. Second, methodological variations across the included studies may have influenced the observed effects of AI. Third, this study focused on five primary moderator variables and did not consider other potentially relevant factors, such as learner characteristics, instructional design, and broader cultural contexts. Therefore, future research should incorporate a larger and more diverse body of studies, including cross-national evidence, to validate and extend the generalisability of these findings. Future studies should also investigate additional moderators that may influence the effectiveness of AI in mathematics education and inform the optimisation of its implementation.

CONCLUSION

The findings of this meta-analysis indicate that the implementation of AI has a statistically significant positive effect on mathematics learning, with the overall effect size classified as moderate. Further moderator analyses identified educational level as a statistically significant moderator, with the estimated effectiveness of AI tending to increase at higher educational levels. Mathematical competence was also found to be a significant moderator, with the largest estimated effects observed for Achievement, Spatial, and Reasoning competencies. In contrast, intervention duration, AI type, and mathematical content area did not demonstrate statistically significant moderation effects. These results indicate that, within the evidence examined in this meta-analysis, no statistically significant differences in AI effectiveness were detected across the categories of these three variables.

This study contributes to the growing body of research on artificial intelligence in mathematics education by providing a synthesis of empirical evidence regarding the effectiveness of AI-based interventions. Through a meta-analytic approach, this study estimates the overall effect of AI on mathematics learning and examines several key moderating variables, including educational level, intervention duration, AI type, mathematical competence, and mathematical content area. By identifying how these factors are associated with variations in the estimated effects of AI, the findings provide insights for researchers and educators into the conditions under which AI may support mathematics learning more effectively. The results also offer practical implications for the design and implementation of AI-based learning environments, particularly regarding the alignment of AI integration with the mathematical competencies targeted in instruction and the characteristics of learners.

REFERENCES

Akavova, A., Temirkhanova, Z., & Lorsanova, Z. (2023). Adaptive learning and artificial intelligence in the educational space. *E3S Web of Conferences, 451*, 06011. https://doi.org/10.1051/e3sconf/202345106011

Al-Nawaiseh, S. J., Tabieh, A. A., Maqableh, W. F., Altawalbeh, M., & Ahmad, F. B. (2024). The effectiveness of using educational robots in enhancing engineering mathematics skills among students in basic school. *International Journal of Education and Practice*, *12*(3), 906-921. https://doi.org/10.18488/61.v12i3.3768

Awang, L. A., Yusop, F. D., & Danaee, M. (2025). Current practices and future direction of

artificial intelligence in mathematics education: A systematic review. *International Electronic Journal of Mathematics Education*, *20*(2), em0823. https://doi.org/10.29333/iejme/16006

Baker, R., & Siemens, G. (2014). Educational Data Mining and Learning Analytics. In R. K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (pp. 253–272). chapter, Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139519526.016

Baker, T., Smith, L., & Anissa, N. (2019). Educ-AI-tion rebooted? exploring the future of artificial intelligence in schools and colleges. In *www.nesta.org.uk*. Nesta. https://media.nesta.org.uk/documents/Future_of_AI_and_education_v5_WEB.pdf

Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., & Tanaka, F. (2018). Social robots for education: A review. *Science Robotics*, *3*(21), 1–10. https://doi.org/10.1126/scirobotics.aat5954

Berida, O., Oguntimilehin, A., Ajisafe, B., Adebusola, T.-O., Babalola, G., Akinduyite, O., Okebule, T., Obamiyi, S., & Sanya, O. (2024, April). SMat-Bot: An interactive chatbot for facilitating students' comprehension of mathematics. In *2024 International Conference on Science, Engineering and Business for Driving Sustainable Development Goals (SEB4SDG)* (pp. 1-8). IEEE. https://doi.org/10.1109/SEB4SDG60871.2024.10630005

Borenstein, M., Hedges, L. V, Higgins, J. P. T., & Rothstein, H. R. (2021). *Introduction to Meta-Analysis. Second Edition*. Wiley. https://doi.org/10.1002/9780470743386

Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. *IEEE Access*, *8*, 75264–75278. https://doi.org/10.1109/ACCESS.2020.2988510

Choi, Y., & McClenen, C. (2020). Development of adaptive formative assessment system using computerized adaptive testing and dynamic bayesian networks. *Applied Sciences (Switzerland)*, *10*(22), 1–17. https://doi.org/10.3390/app10228196

D'Mello, S., & Graesser, A. C. (2015). Feeling, thinking, and computing with affect-aware

learning technologies. Feeling, thinking, and computing with affect-aware learning technologies. In R. Calvo, S. D'Mello, J. Gratch, & A. Kappas (Eds.), *The Oxford handbook of affective computing*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199942237.013.032

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*(2), 455–463. https://doi.org/10.1111/j.0006-341X.2000.00455.x

Ellington, L. M. (2026). Adaptive learning systems: revolutionizing student-centered education. In V. Wang (Ed.), *AI, Personalization, Equity, and the Future of Learning* (pp. 345-368). IGI Global Scientific Publishing. https://doi.org/10.4018/979-8-3373-5097-4.ch012

Fardian, D., Suryadi, D., Prabawanto, S., & Jupri, A. (2025). Integrating Chat-GPT in the classroom: A study on linear algebra learning in higher education. *International Journal of Information and Education Technology*, *15*(4), 732–751. https://doi.org/10.18178/ijiet.2025.15.4.2279

Garcia, M. B., Acut, D. P., Arif, Y. M., Maher, D., Alex, J. K., Wardat, Y., Chiu, T. K. F., & Damaševičius, R. (2026). Modernizing mathematics education with artificial intelligence: A narrative review of ai-powered tools, thematic trends, and instructional applications. In A. Mazı (Ed.), *The Convergence of Mathematics and AI: A New Paradigm in Education* (pp. 153-186). IGI Global Scientific Publishing. https://doi.org/10.4018/979-8-3373-1998-8.ch006

Gürefe, N., Sarpkaya Aktaş, G., & Öksüz, H. (2024). Investigating the Impact of the AI-Supported 5E (AI-s5E) Instructional Model on Spatial Ability. *Behavioral Sciences*, *14*, 682. https://doi.org/10.3390/bs14080682

Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning*. Center for Curriculum Redesign.

Hwang, S. (2022). Examining the effects of artificial intelligence on elementary students' mathematics achievement: A meta-analysis. *Sustainability*, *14*(20), 13185.

https://doi.org/10.3390/su142013185

Khairunnisa, K., Sari, F. F., Anggelena, M., Agustina, D., & Nursa'adah, E. (2022). Penggunaan Effect Size Sebagai Mediasi dalam Koreksi Efek Suatu Penelitian. *Jurnal Pendidikan Matematika (Judika Education)*, *5*(2), 138–151. https://doi.org/10.31539/judika.v5i2.4802

Lampropoulos, G., & Papadakis, S. (2024). Maker education and educational robotics in the age of artificial intelligence. In S. Papadakis & G. Lampropoulos (Eds.), *Intelligent educational robots: Toward personalized learning environments* (pp. 1–8). De Gruyter. https://doi.org/10.1515/9783111352695-001

Létourneau, A., Deslandes Martineau, M., Charland, P., Karran, J. A., Boasen, J., & Léger, P. M. (2025). A systematic review of AI-driven intelligent tutoring systems (ITS) in K-12 education. *Npj Science of Learning*, *10*(1), 29. https://doi.org/10.1038/s41539-025-00320-7

Lu, Y., Pian, Y., Chen, P., Meng, Q., & Cao, Y. (2021). RadarMath: An Intelligent Tutoring System for Math Education. *Proceedings of the AAAI Conference on Artificial Intelligence*, *35*(18), 16087–16090. https://doi.org/10.1609/aaai.v35i18.18020

Memiş, Y. (2025). Examining the potential and pitfalls of AI in problem solving. *Journal of Pedagogical Research*, *9*(2), 177–190. https://doi.org/10.33902/JPR.202532099

Niño, F., Lancheros-Cuesta, D., Jiménez-Valderrama, M., Mestre, G., & Gómez, S. (2023). Systematic Review: Trends in Intelligent Tutoring Systems in Mathematics Teaching and Learning. *International Journal of Education in Mathematics, Science and Technology*, *12*, 203–229. https://doi.org/10.46328/ijemst.3189

Onesimus Laia, H., Martadiputra, B. A. P., & Dahlan, J. A. (2025). The Effect of Logical-Mathematical Intelligence on Mathematics Learning and Moderator Analysis: A Meta-Analysis. *Al-Jabar: Jurnal Pendidikan Matematika*, *16*(1), 129–149. https://doi.org/10.24042/ajpm.v16i1.25539

Opesemowo, O. A. G. (2025). Artificial Intelligence in Mathematics Education: The Pros and

Cons. In M. Khosrow-Pour, D.B.A. (Ed.), *Encyclopedia of Information Science and Technology, Sixth Edition*. Advance online publication. https://doi.org/10.4018/978-1-6684-7366-5.ch084

Ouyang, F., & Xu, W. (2024). The effects of educational robotics in STEM education: A multilevel meta-analysis. *International Journal of STEM Education*, *11*(1). https://doi.org/10.1186/s40594-024-00469-4

Ruppar, T. (2020). Meta-analysis: How to quantify and explain heterogeneity? *European Journal of Cardiovascular Nursing*, *19*(7), 646–652. https://doi.org/10.1177/1474515120944014

Seban, P., & Urban, K. (2024). Examining the Utilisation of Learning Techniques and Strategies Among Pedagogy Students : Implications for Self-Regulated Learning. *Journal of Pedagogy*, *15*, 27–49. https://doi.org/10.2478/jped-2024-0002

Sedgwick, P. (2015). Meta-analyses: What is heterogeneity? *BMJ*, *350*, h1435. https://doi.org/10.1136/bmj.h1435

Sterne, J. A. C., Sutton, A. J., Ioannidis, J. P. A., Terrin, N., Jones, D. R., Lau, J., Carpenter, J., Rücker, G., Harbord, R. M., Schmid, C. H., Tetzlaff, J., Deeks, J. J., Peters, J., Macaskill, P., Schwarzer, G., Duval, S., Altman, D. G., Moher, D., & Higgins, J. P. T. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ, 343*, d4002. https://doi.org/10.1136/bmj.d4002

Surbakti, R., Umboh, S., Pong, M., & Dara, S. (2024). Cognitive Load Theory: Implications for Instructional Design in Digital Classrooms. *International Journal of Educational Narratives*, *2*, 483–493. https://doi.org/10.70177/ijen.v2i6.1659

Sweller, J. (2011). Cognitive load theory. In J. P. Mestre & B. H. Ross (Eds.), *Psychology of learning and motivation* (Vol. 55, pp. 37–76). Academic Press. https://doi.org/10.1016/B978-0-12-387691-1.00002-8

Thalheimer, W., & Cook, S. (2002). How to calculate effect sizes from published research.

Work-Learning Research, *1*(August), 1–9. www.work-learning.com

Wahba, F., Ajlouni, A., & Abumosa, M. (2024). The impact of ChatGPT-based learning statistics on undergraduates' statistical reasoning and attitudes toward statistics. *Eurasia Journal of Mathematics, Science and Technology Education*, *20*, em2468. https://doi.org/10.29333/ejmste/14726

Wu, X. -Y. (2024). Artificial intelligence in L2 learning: A meta-analysis of contextual, instructional, and social-emotional moderators. *System, 126*, 103498. https://doi.org/10.1016/j.system.2024.103498

Yi, L., Liu, D., Jiang, T., & Xian, Y. (2025). The effectiveness of AI on K-12 students' mathematics learning: A systematic review and meta-analysis. *International Journal of Science and Mathematics Education*, *23*(4), 1105–1126. https://doi.org/10.1007/s10763-024-10499-7

Yıldız, S. G., & Körpeoğlu, S. G. (2025). Trends and insights of AI in mathematics education: A bibliometric analysis. *International Electronic Journal of Mathematics Education*, *20*(3), em0837. https://doi.org/10.29333/iejme/16401

Zhang, Y., Zhang, L., Zhang, H., & Wu, X. (2025). Development of a Computerized Adaptive Assessment and Learning System for Mathematical Ability Based on Cognitive Diagnosis. *Journal of Intelligence*, *13*(9), 114. https://doi.org/10.3390/jintelligence13090114

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice*, *41*(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2