International Journal of Advanced and Applied Sciences, 13(7) 2026, Pages: 211-221

Contents lists available at Science-Gate

International Journal of Advanced and Applied Sciences

Journal homepage: https://www.science-gate.com/IJAAS.html

Impact of artificial intelligence integration in STEM education on learners' academic performance: A meta-analytic review

Bakytzhan Kurbanbekov [1], Yerzhan Yedilbayev [1], Serik Polatuly [1,*], Bayan Kuanbayeva [2]

[1]Department of Physics, Khoja Akhmet Yassawi International Kazakh-Turkish University, Turkestan, Kazakhstan
[2]Department of Physics and Technical Disciplines, Khalel Dosmukhamedov Atyrau University, Atyrau, Kazakhstan

ARTICLE INFO

Article history:
Received 31 March 2026
Received in revised form
25 July 2026
Accepted 29 July 2026

Keywords:
AI in education
STEM education
Academic achievement
Meta-analysis
Generative AI

ABSTRACT

The integration of artificial intelligence (AI) technologies into science, technology, engineering, and mathematics (STEM) education has attracted growing scholarly interest due to its potential to improve instructional quality and academic outcomes. This study quantitatively evaluated the effect of AI integration in STEM education on learners' academic achievement using a meta-analytic approach. Following PRISMA guidelines, empirical studies published between 2015 and 2026 were retrieved from Web of Science, IEEE Xplore, and Google Scholar. Studies were included if they used an experimental or quasi-experimental design, included a control group, and provided sufficient statistical information for effect size calculation. A total of 15 studies were included. Effect sizes were calculated using Hedges' g and synthesized using a random-effects model. The results showed a statistically significant positive effect of AI-based STEM interventions on academic achievement ($g = 1.57$, 95% CI [0.985, 2.156], $Z = 5.256$, $p < 0.001$). Moderate-to-substantial heterogeneity was observed ($Q = 72.467$, $p < 0.001$; $I^2 = 73\%$), indicating variation across educational contexts and implementation conditions. Overall, the findings suggest that pedagogically informed AI integration can improve academic achievement in STEM education and support evidence-based educational practice and policy.

1. Introduction

Artificial intelligence (AI) is increasingly integrated into STEM education, offering significant opportunities to enhance instructional effectiveness, personalize learning, and improve academic outcomes (Chen et al., 2020; Xu and Ouyang, 2022). AI technologies, including intelligent tutoring systems, adaptive learning platforms, and educational chatbots, enable data-driven instruction and individualized feedback (Zawacki-Richter et al., 2019). As a key domain for developing advanced cognitive and technological competencies, STEM education particularly benefits from AI integration, which can strengthen learner engagement and support more effective learning processes. Accordingly, examining the effectiveness of AI in STEM education has become a critical direction in contemporary educational research.

Artificial Intelligence in Education (AIEd) encompasses a broad range of technologies, including intelligent tutoring systems, adaptive learning platforms, educational chatbots, learning analytics tools, and machine learning-based predictive systems. Such technologies enable the creation of personalized learning environments by adapting instructional content, feedback mechanisms, and learning strategies to each learner's individual characteristics (Zawacki-Richter et al., 2019). Empirical evidence suggests that AI-based instructional systems improve academic achievement by providing immediate feedback, scaffolding complex tasks, and fostering self-regulated learning (VanLEHN, 2011). Furthermore, by enhancing motivation and sustaining learner interest, both of which constitute critical determinants of academic achievement in STEM disciplines, these systems facilitate improved learning outcomes (Huang et al., 2023).

The integration of AI into STEM education is particularly significant, as STEM subjects inherently demand engagement with complex concepts, interdisciplinary connections, and practical

* Corresponding Author.
Email Address: serik.polatuly@ayu.edu.kz (S. Polatuly)
https://doi.org/10.21833/ijaas.2026.07.020
Corresponding author's ORCID profile:
https://orcid.org/0000-0001-6670-2679

application. AI-based simulations, virtual laboratories, robotics, and mixed reality technologies enable learners to visualize abstract concepts and apply theoretical knowledge in authentic contexts, thereby deepening conceptual understanding and promoting long-term knowledge retention (Mystakidis et al., 2022). Empirical studies demonstrate that the use of AI tools in STEM education improves learners' academic achievement, problem-solving skills, and higher-order thinking abilities (Hwang and Tu, 2021). Nevertheless, the magnitude of effects reported across studies varies considerably, suggesting the potential influence of moderating factors such as educational level, the type of technology employed, the duration of the intervention, and the pedagogical design (Ouyang and Jiao, 2021).

The number of systematic reviews and meta-analyses examining the impact of AI on education has grown substantially in recent years. A considerable body of research has focused on general educational contexts or on specific technologies — such as intelligent tutoring systems or chatbots — reporting moderate positive effects on learning outcomes (Tlili et al., 2025; Wu and Yu, 2024). However, studies dedicated to a comprehensive quantitative synthesis of AI integration effectiveness specifically within STEM education contexts remain limited. Given the distinctive characteristics of STEM disciplines, findings from broader AIEd research may not fully capture the nature and magnitude of effects in this domain. Moreover, the inconsistencies across empirical findings underscore the need for meta-analytic investigations to establish a more precise understanding of AI's effectiveness in STEM education.

Despite growing interest, empirical findings on the effectiveness of AI in STEM education remain inconsistent (Tlili et al., 2025; Wu and Yu, 2024). While some studies report substantial improvements, others indicate context-dependent effects. Therefore, a meta-analytic synthesis is required to establish a more precise estimate of AI effectiveness in STEM education.

This study aims to determine the overall effect of AI integration in STEM education on learners' academic achievement and to examine potential sources of variability across studies.

2. Literature review

In recent years, the application of artificial intelligence in education has expanded rapidly, increasingly recognized as a significant instrument for personalizing instruction, advancing learning analytics, and predicting learners' academic outcomes. Recent research emphasizes the role of artificial intelligence in enabling personalized and adaptive learning environments. Intelligent tutoring systems and adaptive platforms have been shown to improve academic outcomes by tailoring instruction and providing immediate feedback (Bhutoria, 2022). However, their effectiveness depends on instructional design and implementation quality.

In STEM education, AI-supported tools such as simulations, virtual laboratories, and robotics environments facilitate the understanding of complex concepts and enhance problem-solving skills (Chng et al., 2023; Kaushik et al., 2021). Nevertheless, reported effect sizes vary considerably across studies, suggesting the influence of contextual and methodological factors (Ouyang and Jiao, 2021).

The application of AI in STEM education is particularly significant, given that STEM disciplines inherently require complex cognitive processing, abstract reasoning, and practical application. AI-based simulations, virtual laboratories, and robotics tools have been found to enhance learners' conceptual understanding (Chng et al., 2023). Furthermore, research indicates that the use of technology in STEM instruction supports the development of problem-solving capacity and critical thinking skills (Kaushik et al., 2021). Nevertheless, the variability in reported effect sizes suggests that the effectiveness of technology depends on the characteristics of the learning environment, subject content, and instructional strategy. This underscores the need to account for disciplinary specificity and instructional design when implementing AI in STEM education.

Intelligent tutoring systems are among the most extensively studied applications of AI in education. Such systems have been shown to achieve superior outcomes compared to traditional instruction by analyzing learner behavior and delivering immediate, adaptive feedback (Lampropoulos, 2025). A number of meta-analytic studies provide evidence that intelligent tutoring systems have a moderate to large positive effect on learning outcomes (Niño-Rojas et al., 2024). However, effectiveness has been found to vary as a function of subject domain, duration of use, and instructional design, indicating that the efficacy of such systems is context-dependent rather than universal.

More recently, educational chatbots and generative AI tools have been actively integrated into STEM education. Research demonstrates that chatbots can improve learners' motivation, self-efficacy, and academic achievement (Lee et al., 2025). Generative AI tools have similarly shown promise in explaining complex concepts and providing individualized learning support (Kasneci et al., 2023). However, several studies caution that excessive reliance on AI tools may foster cognitive dependency and undermine learners' capacity for independent thinking. This reinforces the argument that AI should be used as a pedagogical support rather than a wholesale replacement for human instruction.

The number of meta-analyses evaluating the effect of AI on learning outcomes has grown considerably. Several studies report an overall positive effect of AI technologies on learners' academic achievement (Zheng et al., 2023), while meta-analytic investigations of adaptive learning systems and virtual agents have identified moderate positive effect sizes (Dai et al., 2024). Nevertheless, the majority of these meta-analyses are oriented toward general educational contexts or individual subject areas, with dedicated analyses of AI effectiveness in STEM education remaining insufficient. This gap underscores the need for specialised meta-analytic research to comprehensively evaluate the effectiveness of AI specifically within STEM educational contexts.

3. Methodology

This study was conducted in accordance with the PRISMA guidelines (Page et al., 2021). A systematic literature search was carried out across the Web of Science, IEEE Xplore, and Google Scholar databases to identify peer-reviewed studies published between 2015 and 2026. The search employed combinations of keywords, including "artificial intelligence in education," "AI in STEM education," "intelligent tutoring systems," "adaptive learning," "chatbots," "academic achievement," and "meta-analysis," refined through the application of Boolean operators (AND, OR).

Studies were included if they met the following criteria: an experimental or quasi-experimental research design, an AI-based intervention in a STEM educational context, inclusion of both experimental and control groups, and sufficient quantitative data to permit effect size calculation. Studies were excluded if they were theoretical or review articles, relied exclusively on qualitative methods, or lacked the requisite statistical information.

The study selection process adhered to the PRISMA framework, encompassing the identification, screening, and eligibility assessment stages (Fig. 1). After duplicates were removed and titles and abstracts screened, full-text articles were evaluated against the inclusion criteria. A total of 15 studies were included in the meta-analysis.

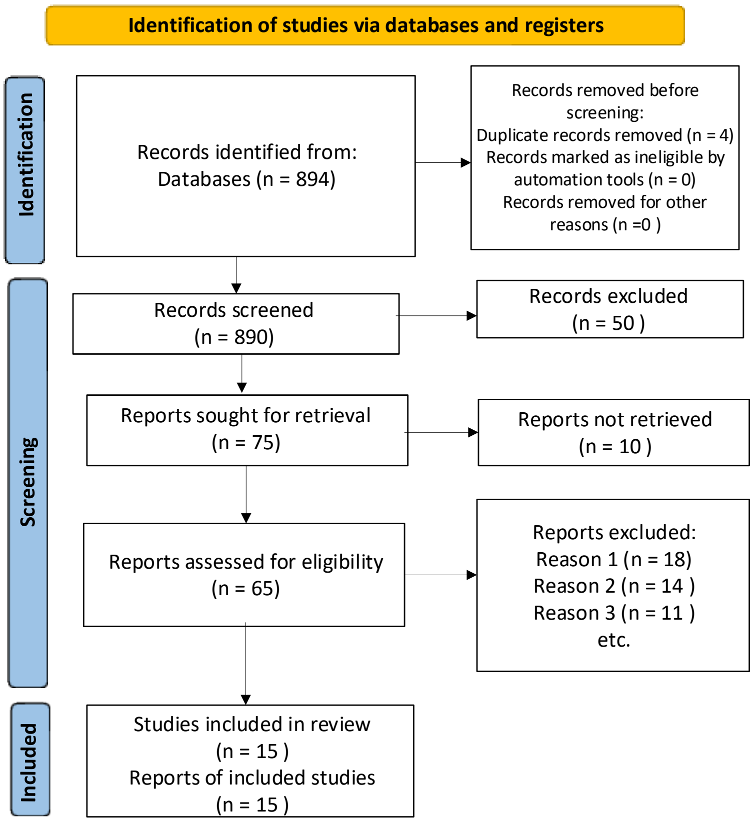

Fig. 1: PRISMA flow diagram of the study selection process

The collected quantitative data were analyzed using the Comprehensive Meta-Analysis (CMA) version 4.0 software. The analysis included the calculation of the overall mean effect size, confidence intervals, the degree of heterogeneity, and relevant statistical indicators, such as the Z-test, Q-test, I^2, τ^2, and related measures.

The quantitative data analysis was conducted based on two key parameters. The first parameter was the effect size expressed as the standardized mean difference, and the second was the standard error estimate. Initially, the standardized mean difference representing the effect size (Hedges' g) was calculated, the formula of which is presented in Eq. 1.

$$g = J \times d \tag{1}$$

where,

$$d = \frac{X_1 - X_2}{S} \tag{2}$$

$$J = 1 - \frac{3}{4(N_1+N_2)-9} \tag{3}$$

$$S = \sqrt{\frac{(n_1-1)S_1{}^2+(n_2-1)S_2{}^2}{n_1+n_2-2}} \tag{4}$$

The standardized error values were calculated using the GraphPad online platform. After determining the effect sizes and standardized errors, the required data were entered into the Comprehensive Meta-Analysis (CMA) version 4.0 software to obtain the final meta-analytical results. In addition, Microsoft Office Excel (MS Excel) 2010 was used to calculate the prediction intervals of the data.

4. Results

A content analysis of the selected studies was conducted to examine the effect of integrating artificial intelligence technologies into STEM education on learners' academic achievement. The analysis systematically reviewed and comparatively assessed the primary findings, pedagogical approaches, and types of technological tools employed across the included studies. This process enabled a comprehensive evaluation of the influence of AI-based instructional interventions on learners' cognitive, academic, and motivational outcomes.

The majority of the reviewed studies found that AI-based STEM learning environments, compared with traditional instructional methods, significantly increase learner engagement, support individualized learning, and improve academic outcomes. In particular, the use of generative AI tools, intelligent tutoring systems, chatbots, and adaptive digital platforms was associated with heightened cognitive engagement and expanded opportunities for learners to construct personalized learning trajectories. Moreover, project-based and problem-based learning environments supported by AI demonstrated clear gains in learners' computational thinking and problem-solving abilities. Generative AI tools and educational chatbots were found to enhance learners' motivation, self-efficacy, and engagement through immediate feedback and personalized support. Intelligent tutoring systems, virtual laboratories, and digital platforms facilitated flexible instructional organization and contributed to sustained improvements in learning outcomes. The mechanisms through which AI technologies influence learners' academic achievement in STEM education are illustrated in Fig. 2.

Studies employing personalised learning systems, intelligent feedback tools, and AI-supported instructional environments consistently identified technological support as a significant factor in enhancing learners' motivation, self-efficacy, and subject mastery. In these works, AI technologies are conceptualised not merely as instructional tools but as pedagogical innovations capable of enhancing the overall effectiveness of the educational process.

Overall, the reviewed studies demonstrate that integrating AI technologies into STEM education is an effective pedagogical approach not only for improving academic outcomes but also for developing 21st-century competencies such as creativity, collaboration, and innovative thinking. These conclusions underscore the relevance of implementing AI-based instructional approaches on a scientifically grounded basis and highlight the need for their quantitative evaluation. Accordingly, a meta-analysis of the selected studies was conducted in the subsequent stage of the research.

In this study, data extracted from the selected scientific literature were structured and systematically presented in Table 1.

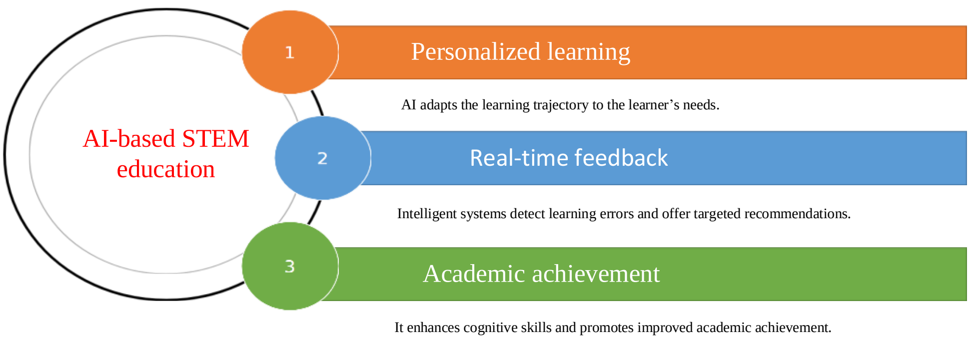

Fig. 2: Mechanisms through which AI technologies influence learners' academic achievement in STEM education

Table 1: Meta-analysis results of the reviewed scientific studies

No.	Study name	Hedges' g	Std. error
1	Liao et al. (2021)	3.27	0.664
2	Kumar et al. (2024)	1.46	0.061
3	Sahito and Khoso (2025)	3.25	1.052
4	Yilmaz and Yilmaz (2023)	0.76	1.059
5	Ayanwale and Omeh (2026)	2.78	1.839
6	Yin et al. (2021)	0.38	0.202
7	Huang and Qiao (2024)	0.35	0.063
8	Omeh and Ayanwale (2025)	2.94	1.828
9	Alneyadi and Wardat (2023)	0.99	1.007
10	Thai et al. (2022)	0.23	2.409
11	del Olmo-Muñoz et al. (2022)	0.56	0.563
12	Lee et al. (2022)	0.76	3.722
13	Essel et al. (2022)	4.53	0.842
14	Kumar (2021)	0.97	0.678
15	Ji et al. (2025)	4.2	1.284

The results of the studies included in the meta-analysis were processed using Comprehensive Meta-Analysis (CMA) version 4.0 software, from which relevant generalized statistical conclusions were drawn. At the initial stage, a comprehensive analysis was conducted of the findings of scientific studies aimed at determining the effect of integrating artificial intelligence technologies into STEM education on learners' academic achievement.

As part of this process, a funnel plot was constructed to assess the potential level of publication bias (Fig. 3). The funnel plot illustrates the distribution of effect sizes from the studies included in the meta-analysis relative to their standard errors and serves as a visual tool for evaluating the degree of symmetry in the research results. This approach enabled a comprehensive assessment of the robustness and reliability of the observed effects of AI-based STEM instructional interventions on learners' academic achievement, as well as an evaluation of the scientific validity of the obtained findings.

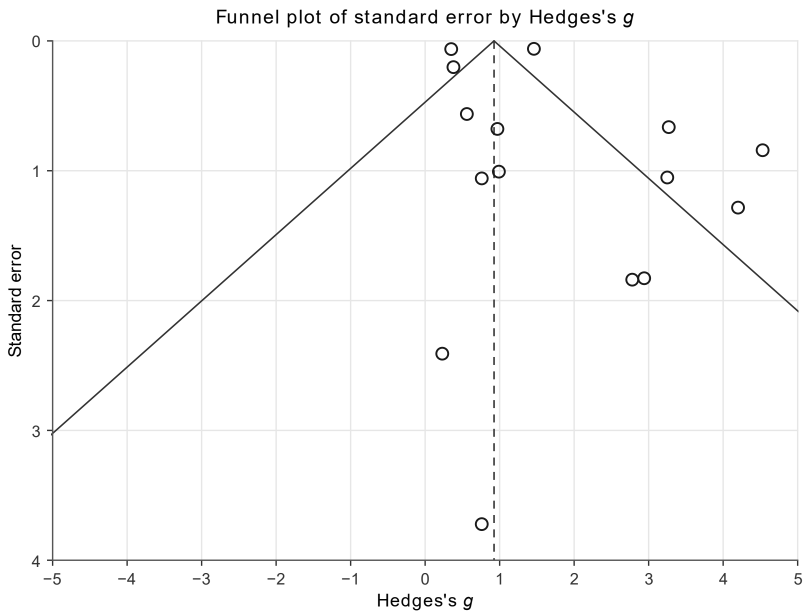

Fig. 3: Funnel plot constructed based on the aggregated scientific data

The potential level of publication bias was assessed using several statistical methods, and the results are presented in Table 2. The Begg and Mazumdar rank correlation test yielded a statistically non-significant result (Kendall's τ = 0.000, p = 0.500), indicating no apparent evidence of publication bias among the studies included in the meta-analysis. Likewise, Egger's regression test did not reveal the presence of a small-study effect (Intercept = 0.987, p = 0.213).

Table 2: Results of the assessment of publication bias

Method	Indicator	*p*-value	Interpretation
Begg & Mazumdar	Kendall's τ = 0	0.5 (1-tailed)	No evidence of publication bias detected
Egger test	Intercept (B_0) = 0.987	0.213 (1-tailed)	No small-study effect detected
Trim & Fill	ES: 1.570 → 0.892	-	Effect size decreased but remained statistically significant
Fail-safe N	790	-	Results are highly robust
Number of studies	k = 15	-	Included in the meta-analysis

Application of the Duval and Tweedie Trim and Fill method demonstrated that the pooled effect size decreased from 1.57 to 0.89 after adjustment for potentially missing studies. Despite this reduction, the adjusted effect size remained positive and statistically meaningful, suggesting that the overall findings are relatively stable and not substantially influenced by publication bias.

Furthermore, Rosenthal's fail-safe N analysis yielded a value of 790, indicating that 790 unpublished null-effect studies would be required to reduce the overall effect to a statistically non-significant level. This number greatly exceeds any plausible estimate of missing studies and therefore supports the robustness of the meta-analytic findings. The results presented in Table 2 suggest that publication bias is either absent or minimal among the included studies. Consequently, the pooled effect size can be considered robust, reliable, and scientifically credible. The distribution of true effects across the included studies is presented in Fig. 4.

The results of the meta-analysis revealed the distribution of true effects describing the variability of effect sizes across the included studies (Fig. 4). The analysis yielded a pooled standardized mean effect size of g = 1.57, with a 95% confidence interval ranging from 0.98 to 2.16. This finding indicates a large positive effect of AI integration in STEM education on learners' academic achievement. Since the confidence interval lies entirely above zero, the overall effect can be considered statistically significant. Furthermore, the 95% prediction interval ranged from −0.26 to 3.40, indicating some variability in effect sizes across the included studies. Although the lower bound of the prediction interval slightly crosses zero, the interval is predominantly located within the positive range. This suggests that, in most comparable educational contexts, AI integration in STEM education is expected to produce positive effects on students' learning outcomes, although the magnitude of the effect may vary depending on instructional conditions and participant characteristics.

Overall, the distribution of true effects demonstrates that AI integration in STEM education represents a statistically supported and educationally meaningful strategy for enhancing learners' academic achievement. The observed effect size indicates substantial benefits compared with traditional instructional approaches.

The meta-analytic findings enabled a comprehensive evaluation of the effectiveness of AI integration in STEM education on learners' academic achievement. To illustrate the magnitude and direction of the effect sizes reported across individual studies, together with their corresponding confidence intervals, the results are presented in the forest plot (Fig. 5).

The individual study estimates reflecting the impact of AI-supported STEM instructional interventions on students' academic achievement, as well as the overall pooled effect size, are presented in Fig. 5.

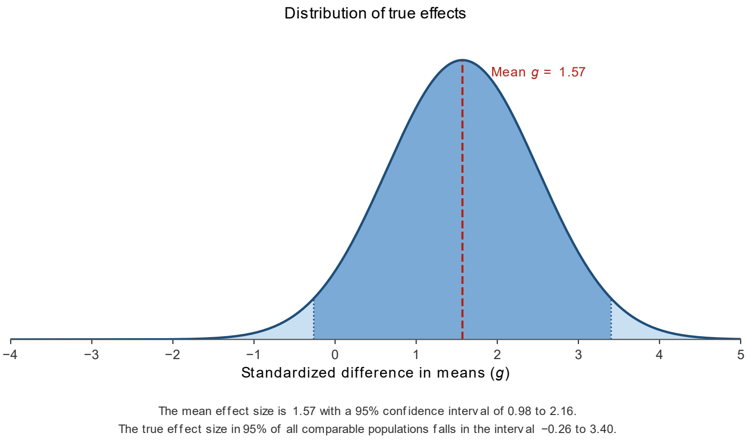

Fig. 4: Distribution of true effects

Study name	Hedges's g	Standard error	Variance	Lower limit	Upper limit	Z-value	p-value
	Statistics for each study						
Liao et al. (2021)	3.270	0.664	0.441	1.969	4.571	4.925	0.000
Kumar et al. (2024)	1.460	0.061	0.004	1.340	1.580	23.934	0.000
Sahito and Khoso (2025)	3.250	1.052	1.107	1.188	5.312	3.089	0.002
Yilmaz and Yilmaz (2023)	0.760	1.059	1.121	-1.316	2.836	0.718	0.473
Ayanwale and Omeh (2026)	2.780	1.839	3.382	-0.824	6.384	1.512	0.131
Yin et al. (2021)	0.380	0.202	0.041	-0.016	0.776	1.881	0.060
Huang and Qiao (2024)	0.350	0.063	0.004	0.227	0.473	5.556	0.000
Omeh and Ayanwale (2025)	2.940	1.828	3.342	-0.643	6.523	1.608	0.108
Alneyadi and Wardat (2023)	0.990	1.007	1.014	-0.984	2.964	0.983	0.326
Thai et al. (2022)	0.230	2.409	5.803	-4.492	4.952	0.095	0.924
del Olmo-Muñoz et al. (2022)	0.560	0.563	0.317	-0.543	1.663	0.995	0.320
Lee et al. (2022)	0.760	3.722	13.853	-6.535	8.055	0.204	0.838
Essel et al. (2022)	4.530	0.842	0.709	2.880	6.180	5.380	0.000
Kumar (2021)	0.970	0.678	0.460	-0.359	2.299	1.431	0.153
Ji et al. (2025)	4.200	1.284	1.649	1.683	6.717	3.271	0.001
Pooled effect (random-effects)	**1.570**	**0.299**	**0.089**	**0.985**	**2.156**	**5.256**	**0.000**
95% prediction interval	*1.570*			*-0.264*	*3.404*		

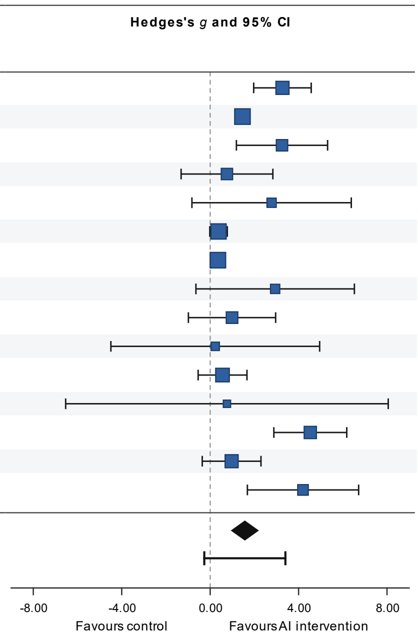

Fig. 5: Graphical representation of the meta-analysis results

4.1. Overview

The present meta-analysis was based on the findings of 15 independent empirical studies evaluating the effect of integrating artificial intelligence technologies into STEM education on learners' academic achievement. Hedges' g was used as the effect size metric because it provides a bias-corrected estimate of the standardized mean difference, particularly appropriate when studies include relatively small sample sizes. The use of a common effect size metric enabled the comparison and synthesis of findings across studies and facilitated the generation of generalized conclusions regarding the effectiveness of AI integration in STEM education.

4.2. Statistical model

A random-effects model was employed throughout the analysis. This model assumes that the included studies represent a random sample drawn from a broader population of potential studies and therefore permits generalisation of the findings to comparable educational contexts. The random-effects model additionally accounts for methodological differences across studies and the natural variability of effect sizes.

4.3. Mean effect size

The pooled mean effect size was g = 1.570, with a 95% confidence interval ranging from 0.985 to 2.156. This finding indicates that the integration of artificial intelligence technologies into STEM education is associated with a statistically significant positive effect on learners' academic achievement. A test of the null hypothesis that the true mean effect size equals zero yielded a statistically significant result ($Z = 5.256$, $p < 0.001$), confirming that the overall effect differs significantly from zero across comparable educational settings.

4.4. Q-test for heterogeneity

Heterogeneity among the included studies was assessed using Cochran's Q statistic. The analysis revealed significant variability in effect sizes across studies ($Q = 72.467$, $df = 14$, $p < 0.001$). This result indicates that the included studies do not share a common effect size and that differences in educational settings, intervention characteristics, participant populations, or implementation procedures may contribute to the observed variation.

4.5. I-squared statistic

The I^2 statistic was 73%, indicating that approximately 73% of the observed variance in effect sizes reflects genuine differences between studies rather than sampling error. This level of heterogeneity can be considered substantial and suggests that the effectiveness of AI integration in STEM education varies across contexts. Therefore, although the overall effect is positive and statistically significant, the magnitude of the effect should be interpreted with appropriate caution.

Overall, the meta-analytic findings indicate that AI-supported STEM educational interventions are associated with improved academic achievement. The pooled effect size suggests a large positive effect; however, the substantial heterogeneity observed across studies indicates that the effectiveness of AI-based instructional approaches may depend on contextual and implementation-related factors. Consequently, the results should be interpreted as evidence of a generally positive association rather than a uniform effect across all educational settings.

To further explore the sources of variability in effect sizes and identify factors potentially influencing the effectiveness of AI interventions, a meta-regression analysis was conducted. The results of the moderator analysis are presented in Table 3.

Table 3: Results of the moderator analysis of the effects of instructional format and intervention duration (random-effects model)

Covariate	Coefficient	Std. error	95% lower	95% upper	Z-value	2-sided P-value
Intercept	2.3464	1.3705	-0.3397	5.0324	1.71	0.0869
Outcome: 2	-0.9625	1.1927	-3.3001	1.3751	-0.81	0.4197
Tool: 2	0.1632	1.0152	-1.8265	2.1529	0.16	0.8723
Approach: 2	-0.8503	1.0712	-2.9498	1.2492	-0.79	0.4273

Test of the model: simultaneous test that all coefficients, excluding the intercept, are zero: $Q = 1.05$, $df = 3$, $p = 0.789$; Goodness-of-fit test: $\tau^2 = 1.3913$, $\tau = 1.1795$, $I^2 = 76.02\%$, $Q = 41.70$, $df = 10$, $p < 0.001$; Comparison with the null model: $\tau^2 = 0.6156$, $\tau = 0.7846$, $I^2 = 73.69\%$, $Q = 72.95$, $df = 13$, $p < 0.001$; Proportion of between-study variance explained by the model: R^2 analog = 0.00 (computed value = −1.26)

Table 3 presents the results of the moderator analysis examining whether outcome type, instructional tool, and instructional approach influenced the overall effect size. The omnibus test of model coefficients was not statistically significant ($Q = 1.05$, $df = 3$, $p = 0.789$), indicating that the set of moderators did not explain a significant proportion of the variance in effect sizes across the included studies. Furthermore, none of the individual moderators reached statistical significance (outcome type: $p = 0.420$; instructional tool: $p = 0.872$; instructional approach: $p = 0.427$), suggesting that variations in outcomes, tools, and instructional approaches were not systematically associated with differences in the effectiveness of AI-supported STEM instruction.

Substantial residual heterogeneity remained after the inclusion of moderator variables ($I^2 = 76.02\%$), and the test for unexplained variance was statistically significant ($Q = 41.70$, $df = 10$, $p < 0.001$), indicating that a considerable proportion of between-study variability was not accounted for by the moderators included in the model. Although the moderator analysis was conducted to explain heterogeneity, the proportion of explained variance was negligible (R^2 analog = 0.00), demonstrating the limited explanatory power of the selected moderator variables.

Overall, these findings suggest that outcome type, instructional tool, and instructional approach were insufficient to explain the heterogeneity observed among the included studies. Therefore, additional factors, such as educational level, learner characteristics, intervention duration, implementation quality, and contextual conditions, should be examined in future research to better understand the sources of variability in the effectiveness of AI-supported STEM instruction.

5. Discussion

The findings of this meta-analysis indicate that the integration of artificial intelligence technologies into STEM education is associated with a statistically significant positive effect on learners' academic achievement. The pooled effect size (g = 1.57) suggests that AI-supported instructional environments may contribute to improved learning outcomes across a range of educational contexts. However, the magnitude of this effect should be interpreted with caution because several included studies reported relatively large effect sizes, and substantial heterogeneity was observed across the dataset.

One possible explanation for the positive association lies in the pedagogical affordances of AI technologies. Intelligent tutoring systems, adaptive learning platforms, educational chatbots, and generative AI tools can provide personalised learning pathways, immediate feedback, and adaptive instructional support. These features may facilitate more effective knowledge acquisition, promote learner engagement, and support deeper conceptual understanding. Such findings are consistent with previous research highlighting the role of technology-enhanced learning environments in improving educational outcomes (Panigrahi et al., 2018).

Furthermore, AI-supported instructional environments may promote self-regulated learning by providing timely feedback and individualised guidance. These characteristics are particularly relevant in STEM education, where learners frequently encounter abstract concepts and complex problem-solving tasks. The ability of AI systems to adapt instructional content to learner needs may therefore contribute to improved academic performance.

Despite the positive overall effect, the analysis revealed substantial heterogeneity among studies (I^2 = 73%). This finding suggests that the effectiveness of AI integration is not uniform across educational settings and may depend on contextual, pedagogical, and methodological factors. Differences in educational level, subject domain, implementation quality, duration of intervention, teacher expertise, and learner characteristics may all contribute to variations in observed outcomes. This interpretation is consistent with Redmond et al. (2018), who argued that educational technologies derive their effectiveness primarily from the quality of pedagogical integration rather than from the technology itself.

The moderator analysis demonstrated that instructional approach, AI tool type, and outcome type did not significantly explain the observed variability in effect sizes. The model accounted for virtually none of the between-study variance ($R^2 \approx 0$), indicating that important sources of heterogeneity remain unidentified. This finding suggests that future research should investigate additional moderators, including implementation fidelity, teacher professional competence, prior student achievement, learning context, and intervention duration. Similar conclusions have been reported by Bao and Koenig (2019), who emphasise the importance of contextual and disciplinary factors in determining educational effectiveness.

Publication bias analyses generally suggested that the findings were relatively robust. However, the Trim-and-Fill procedure reduced the pooled effect size from g = 1.57 to g = 0.89. Although the adjusted effect remained positive, this reduction indicates that some influence of publication bias cannot be completely ruled out. Consequently, the magnitude of the overall effect should be interpreted cautiously.

The findings also support the perspective that AI should not be viewed as a standalone solution capable of automatically improving learning outcomes. Rather, AI functions as a pedagogical tool whose effectiveness depends on alignment with instructional objectives, curriculum design, and established learning theories. In this regard, Galili (2021) emphasised that meaningful learning emerges from the interaction between technological resources, disciplinary knowledge, and educational context.

Overall, the present study provides quantitative evidence that AI-supported STEM education is generally associated with improved academic achievement. At the same time, the substantial heterogeneity, limited explanatory power of the moderator analysis, and potential influence of publication bias highlight the need for cautious interpretation and further high-quality research. Future investigations should focus on identifying the specific conditions under which AI-based interventions are most effective and sustainable across diverse STEM learning environments.

6. Limitations

Several limitations should be considered when interpreting the findings of this study. First, the meta-analysis included only 15 empirical studies, which may limit the generalisability of the results and reduce the statistical power of moderator analyses. Second, substantial heterogeneity remained among the included studies (I^2 = 73%), indicating that important sources of variability may not have been fully captured by the moderator variables examined in this study.

Third, although publication bias analyses did not reveal strong evidence of bias, the Trim-and-Fill procedure reduced the pooled effect size from g = 1.57 to g = 0.89, suggesting that some influence of publication bias cannot be entirely excluded. Fourth, several studies reported relatively large effect sizes, which may have influenced the overall pooled estimate despite the use of Hedges' g and a random-effects model.

Fifth, a formal risk-of-bias assessment using instruments such as MMAT, ROBINS-I, or Cochrane RoB was not conducted. Therefore, variations in methodological quality across studies could not be systematically evaluated. Finally, the moderator analysis explained little of the between-study variance, suggesting that additional factors, including implementation fidelity, teacher expertise, learner characteristics, educational level, and intervention duration, should be examined in future research.

Despite these limitations, the present study provides a quantitative synthesis of current evidence regarding AI integration in STEM education and offers a useful foundation for future research and educational practice.

7. Conclusion

This meta-analysis synthesized the findings of 15 empirical studies to examine the effect of integrating artificial intelligence into STEM education on learners' academic achievement. The results indicate that AI-based instructional approaches exert a statistically significant and educationally meaningful positive effect on learning outcomes.

The findings demonstrate that AI technologies, including intelligent tutoring systems, adaptive learning platforms, and generative AI tools, can enhance academic achievement by providing personalized learning experiences, facilitating cognitive engagement, and supporting the development of higher-order thinking skills.

Nevertheless, the substantial heterogeneity observed across the included studies indicates that the effectiveness of AI integration is highly contingent upon contextual and pedagogical factors. This underscores the importance of designing and implementing AI-supported learning environments aligned with evidence-based instructional strategies, rather than relying solely on the technological capabilities of the tools used.

The present study contributes to the growing body of research on artificial intelligence in education by providing robust quantitative evidence within a STEM-specific context. At the same time, the findings highlight the need for further investigation to identify key moderating variables, including learner characteristics, subject domains, and implementation conditions.

Future research should also prioritize longitudinal and large-scale experimental designs to examine the sustainability and transferability of learning gains over time. Overall, the present study affirms that the pedagogically informed integration of AI is a promising, empirically substantiated approach to enhancing academic achievement in STEM education.

List of abbreviations

AI	Artificial intelligence
AIEd	Artificial intelligence in education
CI	Confidence interval
CMA	Comprehensive meta-analysis
df	Degrees of freedom
ES	Effect size
I^2	I-squared (heterogeneity statistic)
PRISMA	Preferred reporting items for systematic reviews and meta-analyses
Q	Cochran's Q statistic
STEM	Science, technology, engineering, and mathematics
τ	Tau (between-study standard deviation)
τ^2	Tau-squared (between-study variance)
Z-test	Z statistical test

Acknowledgment

This research has been funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant AP26105437). This article was prepared within the framework of the postdoctoral program of Khoja Akhmet Yassawi International Kazakh-Turkish University.

Compliance with ethical standards

Conflict of interest

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

Alneyadi S and Wardat Y (2023). ChatGPT: Revolutionizing student achievement in the electronic magnetism unit for eleventh-grade students in Emirates schools. Contemporary Educational Technology, 15(4): ep448. https://doi.org/10.30935/cedtech/13417

Ayanwale MA and Omeh CB (2026). AI-supported problem-based learning for enhancing computational thinking skills in STEM education. Computers in Human Behavior: Artificial Humans, 7: 100263. https://doi.org/10.1016/j.chbah.2026.100263

Bao L and Koenig K (2019). Physics education research for 21st century learning. Disciplinary and Interdisciplinary Science Education Research, 1: 2. https://doi.org/10.1186/s43031-019-0007-8

Bhutoria A (2022). Personalized education and artificial intelligence in the United States, China, and India: A systematic review using a human-in-the-loop model. Computers and Education: Artificial Intelligence, 3: 100068. https://doi.org/10.1016/j.caeai.2022.100068

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

Chng E, Tan AL, and Tan SC (2023). Examining the use of emerging technologies in schools: A review of artificial intelligence and immersive technologies in STEM education. Journal for STEM Education Research, 6(3): 385-407. https://doi.org/10.1007/s41979-023-00092-y

Dai CP, Ke F, Pan Y, Moon J, and Liu Z (2024). Effects of artificial intelligence-powered virtual agents on learning outcomes in computer-based simulations: A meta-analysis. Educational Psychology Review, 36: 31. https://doi.org/10.1007/s10648-024-09855-4

del Olmo-Muñoz J, González-Calero JA, Diago PD, Arnau D, and Arevalillo-Herráez M (2022). Using intra-task flexibility on an intelligent tutoring system to promote arithmetic problem-solving proficiency. British Journal of Educational Technology, 53(6): 1976-1992. https://doi.org/10.1111/bjet.13228

Essel HB, Vlachopoulos D, Tachie-Menson A, Johnson EE, and Baah PK (2022). The impact of a virtual teaching assistant (chatbot) on students' learning in Ghanaian higher education. International Journal of Educational Technology in Higher Education, 19: 57. https://doi.org/10.1186/s41239-022-00362-6

Galili I (2021). Scientific knowledge as a culture: A paradigm of knowledge representation for the meaningful teaching and learning of science. In: Galili I (Ed.), Scientific knowledge as a culture: The pleasure of understanding: 245–275. Springer, Cham, Switzerland. https://doi.org/10.1007/978-3-030-80201-1

Huang AY, Lu OH, and Yang SJ (2023). Effects of artificial intelligence–enabled personalized recommendations on learners' learning engagement, motivation, and outcomes in a flipped classroom. Computers & Education, 194: 104684. https://doi.org/10.1016/j.compedu.2022.104684

Huang X and Qiao C (2024). Enhancing computational thinking skills through artificial intelligence education at a STEAM high school. Science & Education, 33: 383-403. https://doi.org/10.1007/s11191-022-00392-6

Hwang GJ and Tu YF (2021). Roles and research trends of artificial intelligence in mathematics education: A bibliometric mapping analysis and systematic review. Mathematics, 9(6): 584. https://doi.org/10.3390/math9060584

Ji Y, Zhan Z, Li T, Zou X, and Lyu S (2025). Human–machine cocreation: The effects of ChatGPT on students' learning performance, AI awareness, critical thinking, and cognitive load in a STEM course toward entrepreneurship. IEEE Transactions on Learning Technologies, 18: 402-415. https://doi.org/10.1109/TLT.2025.3554584

Kasneci E, Seßler K, Küchemann S et al. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103: 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kaushik R, Parmar M, and Jhamb S (2021). Roles and research trends of artificial intelligence in mathematics education. In the 2nd International Conference on Computational Methods in Science & Technology (ICCMST), IEEE, Mohali, India: 202-205. https://doi.org/10.1109/ICCMST54943.2021.00050

Kumar JA (2021). Educational chatbots for project-based learning: Investigating learning outcomes for a team-based design course. International Journal of Educational Technology in Higher Education, 18: 65. https://doi.org/10.1186/s41239-021-00302-w **PMid:34926790 PMCid:PMC8670881**

Kumar KS, Selvan T, Mahendraprabu M, Ganesan K, Ramnath R, and Kumar NS (2024). Examining the role of virtual reality, augmented reality, and artificial intelligence in adapting STEM education for next-generation inclusion. International Journal of Emerging Knowledge Studies, 2(12): 876-883. https://doi.org/10.70333/ijeks-02-12-025

Lampropoulos G (2025). Augmented reality, virtual reality, and intelligent tutoring systems in education and training: A systematic literature review. Applied Sciences, 15(6): 3223. https://doi.org/10.3390/app15063223

Lee YF, Hwang GJ, and Chen PY (2022). Impacts of an AI-based *cha*bot on college students' after-class review, academic performance, self-efficacy, learning attitude, and motivation. Educational Technology Research and Development, 70: 1843-1865. https://doi.org/10.1007/s11423-022-10142-8

Lee YF, Hwang GJ, and Chen PY (2025). Technology-based interactive guidance to promote learning performance and self-regulation: A chatbot-assisted self-regulated learning approach. Educational Technology Research and Development, 73: 2279-2304. https://doi.org/10.1007/s11423-025-10478-x

Liao J, Yang J, and Zhang W (2021). The student-centered STEM learning model based on artificial intelligence project: A case study on intelligent car. International Journal of Emerging Technologies in Learning (iJET), 16(21): 100-120. https://doi.org/10.3991/ijet.v16i21.25001

Mystakidis S, Christopoulos A, and Pellas N (2022). A systematic mapping review of augmented reality applications to support STEM learning in higher education. Education and Information Technologies, 27: 1883-1927. https://doi.org/10.1007/s10639-021-10682-1

Niño-Rojas F, Lancheros-Cuesta D, Jiménez-Valderrama MTP, Mestre G, and Gómez S (2024). Systematic review: Trends in intelligent tutoring systems in mathematics teaching and learning. International Journal of Education in Mathematics, Science and Technology, 12(1): 203-229. https://doi.org/10.46328/ijemst.3189

Omeh CB and Ayanwale MA (2025). Artificial intelligence meets PBL: Transforming computer-robotics programming motivation and engagement. Frontiers in Education, 10: 1674320. https://doi.org/10.3389/feduc.2025.1674320

Ouyang F and Jiao P (2021). Artificial intelligence in education: The three paradigms. Computers and Education: Artificial Intelligence, 2: 100020. https://doi.org/10.1016/j.caeai.2021.100020

Page MJ, McKenzie JE, Bossuyt PM et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372: n71. https://doi.org/10.1136/bmj.n71 **PMid:33782057**

Panigrahi R, Srivastava PR, and Sharma D (2018). Online learning: Adoption, continuance, and learning outcome—A review of literature. International Journal of Information Management, 43: 1-14. https://doi.org/10.1016/j.ijinfomgt.2018.05.005

Redmond P, Abawi L, Brown A, Henderson R, and Heffernan A (2018). An online engagement framework for higher education. Online Learning Journal, 22(1): 183-204. https://doi.org/10.24059/olj.v22i1.1175

Sahito ZH and Khoso FJ (2025). The effectiveness of technology-enhanced learning tools, including virtual labs and AI-powered platforms, in improving STEM education outcomes among secondary school students. International Journal of Social Science & Entrepreneurship, 5(4): 28-48. https://doi.org/10.58622/sknm9428

Thai KP, Bang HJ, and Li L (2022). Accelerating early math learning with research-based personalized learning games: A cluster randomized controlled trial. Journal of Research on Educational Effectiveness, 15(1): 28-51. https://doi.org/10.1080/19345747.2021.1969710

Tlili A, Saqer K, Salha S, and Huang R (2025). Investigating the effect of artificial intelligence in education (AIEd) on learning achievement: A meta-analysis and research synthesis. Information Development, 41(3): 825-842. https://doi.org/10.1177/02666669241304407

VanLEHN K (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. Educational Psychologist, 46(4): 197-221. https://doi.org/10.1080/00461520.2011.611369

Wu R and Yu Z (2024). Do AI chatbots improve students learning outcomes? Evidence from a meta-analysis. British Journal of Educational Technology, 55(1): 10-33. https://doi.org/10.1111/bjet.13334

Xu W and Ouyang F (2022). The application of AI technologies in STEM education: A systematic review from 2011 to 2021. International Journal of STEM Education, 9: 59. https://doi.org/10.1186/s40594-022-00377-5

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

Yin J, Goh TT, Yang B, and Xiaobin Y (2021). Conversation technology with micro-learning: The impact of chatbot-based learning on students' learning motivation and performance. Journal of Educational Computing Research, 59(1): 154-177. https://doi.org/10.1177/0735633120952067

Zawacki-Richter O, Marín VI, Bond M, and 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

Zheng L, Niu J, Zhong L, and Gyasi JF (2023). The effectiveness of artificial intelligence on learning achievement and learning perception: A meta-analysis. Interactive Learning Environments, 31(9): 5650-5664. https://doi.org/10.1080/10494820.2021.2015693