JKTP: Jurnal Kajian Teknologi Pendidikan
Vol. 9, No. 4, November 2026: 499-511
e-ISSN: 2615-8787
DOI: 10.17977/um038v9i42026p499-511
https://journal-fip.um.ac.id/index.php/jktp

EFFECTIVENESS OF ARTIFICIAL INTELLIGENCE ON STUDENTS' CRITICAL THINKING SKILLS: A META-ANALYSIS STUDY

Santa Aulia Devi Rachmadhani, Ali Muhson, Risky Setiawan
Penelitian dan Evaluasi Pendidikan, Universitas Negeri Yogyakarta, Indonesia
Corresponding Author E-Mail: santaaulia.2025@student.uny.ac.id

Article History
Received: 08 October 2026, Accepted: 04 November 2026, ***Published: 11 November 2026***

Abstrak
Bukti empiris pengaruh AI terhadap kemampuan berpikir kritis siswa masih terfragmentasi dan belum disintesis secara sistematis. Penelitian ini bertujuan untuk menghitung besarnya efek intervensi berbasis AI terhadap kemampuan berpikir kritis, menguji bias publikasi, serta menelaah peran empat faktor moderator, yaitu wilayah, jenjang pendidikan, lama perlakuan, dan jenis AI yang digunakan. Analisis dibatasi pada studi kuasi-eksperimen yang memiliki kelompok kontrol. Meta-analisis dilakukan terhadap studi yang diambil dari Scopus, ScienceDirect, dan ERIC dengan mengikuti pedoman PRISMA. Ukuran efek dihitung menggunakan Hedges' g dalam model efek acak. Hasil analisis menunjukkan adanya efek besar yang signifikan, disertai heterogenitas yang tinggi, dan tidak ditemukan bias publikasi. Analisis moderator memperlihatkan perbedaan efek yang signifikan pada keempat faktor tersebut. Efek terbesar muncul pada studi di Afrika, jenjang perguruan tinggi, intervensi dengan durasi panjang, serta penggunaan AI campuran. Intervensi berbasis AI cenderung meningkatkan kemampuan berpikir kritis siswa. Namun, besarnya efek bergantung pada pada cara intervensi dirancang dan penerapannya. Dengan demikian, integrasi AI dalam pembelajaran sebaiknya disusun terarah sesuai kerangka pedagogis dan didukung intervensi yang cukup panjang. Temuan ini memperkuat bukti empiris hubungan AI dan berpikir kritis.

Kata Kunci: Kecerdasan Buatan; Berpikir Kritis; Meta-Analisis; Analisis Moderator; Ukuran Efek

Abstract
Empirical evidence on the effect of artificial intelligence (AI) on students' critical thinking skills remains fragmented and has not been systematically synthesized. This study aims to quantify the effect of AI-based interventions on critical thinking skills, test for publication bias, and examine four moderators: region, education level, treatment duration, and type of AI used. The analysis was limited to quasi-experimental studies with a control group. A meta-analysis was conducted on studies retrieved from Scopus, ScienceDirect, and ERIC, following the PRISMA guidelines. Effect sizes were calculated using Hedges' g under a random-effects model. The results show a significant large effect with high heterogeneity, and no evidence of publication bias was found. Moderator analyses revealed significant differences in effect size across all four factors. The largest effects were observed in studies conducted in Africa, at the higher education level, with long-duration interventions, and with mixed AI use. Overall, AI-based interventions tend to improve students' critical thinking skills. However, the magnitude of the effect depends on how the intervention is designed and implemented. Therefore, AI integration in learning should be purposefully structured according to pedagogical frameworks and supported by sufficiently long interventions. These findings strengthen the empirical evidence on the relationship between AI and critical thinking.

Keywords: *Artificial Intelligence; Critical Thinking; Meta-Analysis; Moderator Analysis; Effect Size*

To cite this article:
Santa, A. D. R., Muhson, A., & Setiawan, R. (2026). Effectiveness of artificial intelligence on students' critical thinking skills: A meta-analysis study. *JKTP: Jurnal Kajian Teknologi Pendidikan*, *9*(4), 499–511. doi: 10.17977/um038v9i42026p499-511

INTRODUCTION

In an era of abundant digital information and rapid technological change, critical thinking has become one of the most important competencies for individuals and society (García-Carmona, 2025). Without it, a person can easily be swept along by misinformation or fail to make sound decisions amid abundant yet unfiltered information. Memorization and passive content mastery are no longer sufficient. Learners must be able to test assumptions, evaluate evidence, build coherent arguments, and face new unstructured problems (Andrian et al., 2025; Najah et al., 2024). Studies in Indonesian classrooms likewise stress that critical thinking needs to be developed from primary education through vocational high school (Arimbawa et al., 2025; Linsih et al., 2025).

Critical thinking, as defined by Facione (2011), encompasses six skills: interpretation, analysis, evaluation, inference, explanation, and self-regulation. These skills involve interactions between cognitive and metacognitive processes, which makes them essential for evaluating information sources, detecting bias, and formulating logical arguments. Partnership for 21st Century Skills (2015) places critical thinking among four core competencies (4Cs), alongside communication, collaboration, and creativity, while teacher support, distance learning platforms, and socioeconomic status influence how far learners can develop it (İdil et al., 2024)..

Developing these skills requires instructional approaches that go beyond factual mastery and actively engage learners in analysis, evaluation, and reflection. Technology-based innovations that can support such engagement have attracted growing attention, and artificial intelligence (AI) is among the most relevant to examine. Modern AI systems, which encompass machine learning algorithms, artificial neural networks, and natural language processing, are designed to process large-scale data, recognize patterns, and provide contextually relevant recommendations (T. Li et al., 2025; Yas et al., 2025). In education, AI enables predictions of student performance and more responsive feedback than conventional approaches (Maia et al., 2023). Unlike static modules, AI systems can identify each student's specific difficulties and adapt learning pathways automatically, although their use must consider algorithmic bias and data privacy (Alghamdi & Alghizzi, 2025).

Theoretically, the mechanisms by which AI supports critical thinking can be explained through learning theories. From Vygotsky (1978) sociocultural perspective, AI can be viewed as an adaptive scaffolding agent within the Zone of Proximal Development (ZPD) that helps learners surpass their independent capabilities. From a constructivist perspective (Piaget, 1970; Vygotsky, 1978), AI can act as a facilitator of active knowledge construction. Both perspectives assume that AI is used to promote learners' thinking rather than replace it.

Without appropriate guidance, however, AI can also pose risks. Cognitive offloading, the tendency to delegate thinking effort to AI systems, may impede higher-order thinking (Gerlich, 2025; Jose et al., 2025). Tian and Zhang (2025) showed that dependence on AI correlates with cognitive fatigue and reduced analytical independence, Abbas et al. (2025) found that excessive dependence is associated with lower academic achievement, and Zhai et al. (2024) warned that AI dialogue systems risk promoting uncritical acceptance of outputs. In addition, AI hallucination, the tendency to generate convincing but inaccurate content, requires learners to evaluate AI outputs critically (Kamel, 2024; Y. Li et al., 2025; Y. Sun et al., 2024).

Empirical findings are therefore mixed. Essel et al. (2024) reported positive effects of ChatGPT on undergraduate students' cognitive skills when its use was accompanied by explicit instructional structure, and Cao & Gao (2026) found positive effects of a pedagogically customized machine translation tool on critical thinking in foreign language education. A quasi-experimental study in Indonesian higher education also found that AI use significantly increased students' creativity in writing scientific papers (Sinaga et al., 2025). In contrast, Mai, (2026) reported negative effects of

ChatGPT on EFL students' argumentative skills when the chatbot was used without additional scaffolding, and Bueno-Picazo & Tirado-Olivares (2026) found a small negative effect of generative AI in geography education. These differences largely reflect how AI is integrated, as a thinking tool or as a substitute for thinking, which suggests that effectiveness depends on the quality of the accompanying instructional design rather than on the technology itself.

Meta-analysis can reconcile such inconsistent findings by statistically aggregating effect sizes across studies, producing estimates that are more precise than any single study (Borenstein et al., 2009b; Glass, 1976). Several recent meta-analyses have examined the effects of AI on critical thinking and higher-order thinking. Sun et al. (2026) conducted a three-level meta-analysis of generative AI tools and critical thinking, Zhao et al. (2025) examined generative AI and higher-order thinking with critical thinking as one of its domains, and Ilgun Dibek et al., (2025) synthesized the influence of AI tools on higher-order thinking skills more broadly. Nevertheless, more diverse AI-based interventions, such as chatbots, adaptive learning systems, intelligent tutoring systems, and combinations of AI types, have rarely been analyzed as a group in relation to critical thinking. In addition, evidence from controlled quasi-experimental studies has seldom been synthesized with region, educational level, treatment duration, and type of AI considered simultaneously as moderators.

This study differs from previous syntheses in three ways. It covers various AI-based interventions, such as chatbots, generative AI, adaptive learning, intelligent tutoring systems, and mixed AI. It also restricts the evidence to quasi-experimental studies with control groups published between 2022 and 2026. In addition, it examines region, educational level, treatment duration, and type of AI as moderators and assesses publication bias through funnel plot and Egger's test. This study therefore aims to estimate the overall effect of AI-based interventions on critical thinking under a random-effects model, to examine the roles of the four moderators in explaining variation in that effect, and to assess the extent to which publication bias affects the estimate.

METHOD

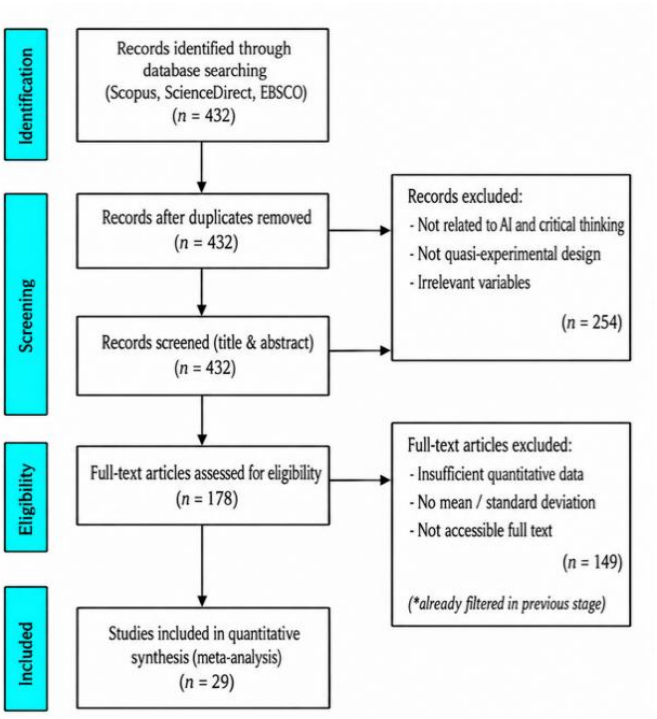


Figure 1. PRISMA Flow Diagram of Study Selection

This study used a meta-analysis design to synthesize empirical evidence on the effect of artificial intelligence (AI)-based interventions on students' critical thinking skills. The procedure followed the PRISMA 2020 guidelines (Page et al., 2021) and recommended meta-analytic methodology (Borenstein, Hedges, et al., 2009; Higgins et al., 2003). All calculations were performed in RStudio using the metafor package.

The data were primary studies searched systematically in Scopus, ScienceDirect, and ERIC for publications from January 2022 to March 2026. Keywords combined AI-related terms ("artificial intelligence", "generative AI", "ChatGPT", "AI chatbot", "intelligent tutoring", "adaptive learning") with outcome terms ("critical thinking", "higher-order thinking", "analytical thinking") using the Boolean operators AND and OR. Of the 432 articles identified, 178 were assessed at the full-text stage and 29 studies met the criteria, as shown in Figure 1. Studies were included if they (1) used a quasi-experimental design with a control group; (2) implemented an AI-based intervention; (3) measured critical thinking with a validated instrument; (4) reported means, standard deviations, and sample sizes; (5) were published in an indexed journal; and (6) were written in English.

The data collection instrument was a coding sheet designed in advance to keep categorization consistent. Coding was carried out by two coders, and disagreements were resolved through discussion. In addition to statistical data, each study was coded on four moderator variables, namely country/region, educational level, treatment duration, and type of AI, following the scheme in Table 1. These variables were chosen because they are theoretically relevant to the effectiveness of AI interventions in education.

Tabel 1. Code of Moderator

Variable	Code
Country/Region	1 = Asia
	2 = Europe
	3 = Africa
Educational Level	1 = Primary School
	2 = Senior High School
	3 = Higher Education
	4 = Mixed
Treatment Duration	1 = ≤ 4 weeks
	2 = 5–8 weeks
	3 = > 8 weeks
Type of AI	1 = AI Chatbot
	2 = Adaptive Learning
	3 = Generative AI
	4 = Mixed AI
	5 = Other

Note. Categories without studies (America, Mixed region, Junior High School, and Mixed level) were not analyzed. Subgroups with k < 2 were excluded because a pooled effect and between-study variance cannot be estimated.

Each study was represented by one effect size, Hedges's g, the standardized mean difference between the experimental (E) and control (C) groups with correction factor J for small-sample bias. This metric was chosen because it is more conservative than Cohen's d for small samples. The formula is as follows.

$$g = J(df)x\ (M_E - M_C)/S_{pooled} \qquad (1)$$

In this formula, S_{pooled} is the pooled standard deviation and J(df) = 1 − 3/(4df − 1). Variances and standard errors were calculated following (Hedges & Olkin, 1985). Studies reporting only t, F,

or d statistics were converted using the formulas of (Borenstein, Hedges, et al., 2009). Effect sizes were interpreted following Cohen (1992): approximately 0.2 small, 0.5 medium, and 0.8 large.

Heterogeneity was tested with Cochran's Q and the I^2 index, classified according Higgins et al. (2003): < 25% low, 25–50% moderate, 50–75% moderate-high, and > 75% high. I^2 was calculated in two ways: under the fixed-effect model with the Higgins formula, (Q − df)/Q, and under the random-effects model with the REML estimate, $\tau^2/(\tau^2 + \bar{\sigma}^2)$, where τ^2 is the between-study variance. When heterogeneity was high, a random-effects model with the REML estimator was used as the main model and a fixed-effect model as a comparison. Publication bias was examined through visual inspection of the funnel plot and Egger's regression test (Egger et al., 1997), with $p < 0.05$ indicating significant bias. Moderator analysis used a subgroup approach for the four variables above; the effect size of each subgroup was estimated separately and differences were tested with the omnibus Q-between test (Borenstein, Hedges, et al., 2009). Subgroups with $k < 2$ were excluded from the analysis.

RESULT

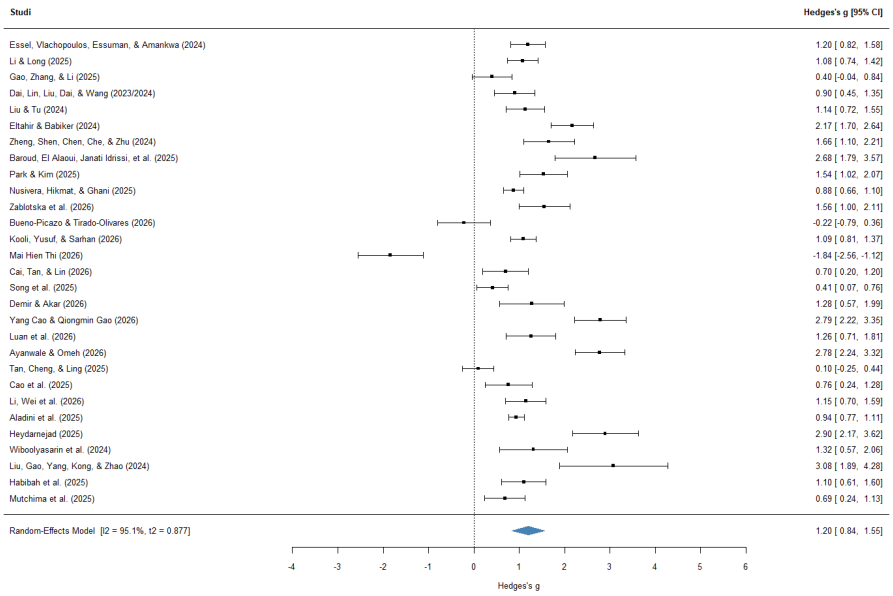


Figure 2. Forest Plot of Overall Effect Size

The random-effects model (REML) yielded an overall effect size of g = 1.1971 (95% CI [0.8416; 1.5527]; Z = 6.5985; $p < 0.001$). This g value falls in the large category according to Cohen (1992). For comparison, the fixed-effects model yielded g = 1.0206 (95% CI [0.9447; 1.0966]) with a narrower CI range. The difference in CI width between the two models reflects between-study variance ($\tau^2 = 0.8775$; $\tau = 0.9367$. The results of both models are shown in Table 2.

Table 2. Heterogeneity Test Results and Effect Size Estimates

Model	Number of Studies	Effect Size	95% Confidence Interval		Heterogeneity			
			Lower Limit	Upper Limit	Q-Value	df	*p-Value*	I-Squared
Fixed	29	1.0206	0.9447	1.0966	300.064	28	0.000	90.67%
Random	29	1.1971	0.8416	1.5527				95,08%

Note. k = 29 studies. Fixed-effects I^2 calculated using the Higgins formula: (Q − df)/Q = 90.67%. Random-effects I^2 is the REML estimate: $\tau^2/(\tau^2 + \bar{\sigma}^2)$ = 95.08%.

The Cochran's Q test yielded Q = 300.064 with df = 28 and p < 0.001. This very small p value rejects the null hypothesis of homogeneity. The I^2 index for the random-effects model reached 95.08%. According to the criteria of (Higgins et al., 2003), this value falls in the high category (> 75%). Therefore, the random-effects model is the appropriate choice, and moderator analyses are needed to identify factors that explain this heterogeneity. Publication bias was evaluated through visual inspection of the funnel plot and Egger's regression test (Figure 3).

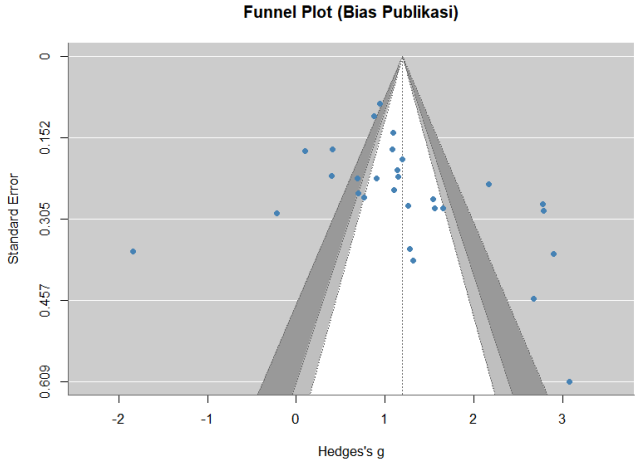


Figure 3. Funnel Plot of Standard Error by Hedges's g

The funnel plot showed a relatively symmetric distribution of study points around the mean effect value, although a few studies were located outside the 95% confidence contour in the lower-right region. Egger's regression test yielded t = 1.576 (df = 27; p = 0.127), a value not significant at α = 0.05. The overall effect size estimate can be considered stable and undistorted by potentially unpublished studies.

Table 3. Test of the moderating effect

Moderator	Subgroup	k	ES	95% CI		Two-TailedTest		Intergroup Effect
				Lower Limit	Upper Limit	Z-Value	p-Value	
Country/ Region	Asia	23	1,121	0,731	1,511	5,630	<0.001	Q = 47.346 p < 0.001
	Europe	2	0,672	-1,068	2,412	0,757	0,449	
	Africa	4	1,887	0,984	2,790	4,097	<0.001	
Educational Level	Primary School	3	0,645	-0,222	1,513	1,458	0,145	Q = 43.708 p < 0.001
	Senior High School	6	1,068	-0,275	2,411	1,559	0,119	
	Higher Education	20	1,305	0,970	1,640	7,641	<0.001	
Treatment Duration	≤4 weeks	4	1,118	-0,019	2,254	1,927	0,054	Q = 51.402 p < 0.001
	5–8 weeks	6	0,495	-0,427	1,417	1,051	0,293	
	>8 weeks	19	1,426	1,051	1,801	7,453	<0.001	
Type of AI	AI Chatbot	11	0,986	0,244	1,729	2,604	0,009	Q = 43.851 p < 0.001
	Generative AI	5	0,865	0,591	1,139	6,183	<0.001	
	Mixed AI	5	1,640	0,751	2,528	3,618	<0.001	
	Other	7	1,473	0,784	2,162	4,192	<0.001	

Note. k = number of studies; ES = Hedges's g; CI = confidence interval. Two-tailed p values; α = 0.05.

Country/region served as a significant moderator (Q-between = 47.346; df = 2; p < 0.001). The Africa subgroup (k = 4) yielded the largest effect size with g = 1.887. The Asia subgroup, with the largest number of studies (k = 23), yielded g = 1.121. The Europe subgroup (k = 2) was not statistically significant (g = 0.672). This non-significant result more likely reflects limitations in the number of studies than AI's ineffectiveness in Europe.

The educational level moderator test yielded Q-between = 43.708 (df = 2; $p < 0.001$). The higher education subgroup (k = 20) obtained g = 1.305 and was the only statistically significant subgroup. Studies at this level include Li & Long (2025), Gao et al., (2025), Park & Kim (2025), Nusivera et al. (2025), Cai et al. (2026), and Luan et al. (2026), along with several other studies. The senior high school subgroup (k = 6) yielded g = 1.068, while the primary school subgroup (k = 3) yielded g = 0.645. Both subgroups had relatively large positive effect sizes but did not reach statistical significance.

Treatment duration was the moderator with the largest Q-between among the four tested (Q = 51.402; df = 2; $p < 0.001$). Interventions lasting more than 8 weeks (k = 19) yielded the largest and most consistent effect size with g = 1.426. Studies in this subgroup include Liu & Tu (2024), Wiboolyasarin et al. (2024), Cao & Gao (2026), and Luan et al. (2026). Short interventions of ≤4 weeks (k = 4) yielded g = 1.118, very close to the significance threshold. The 5–8 week duration (k = 6) actually yielded the lowest effect size, g = 0.495, and was not significant.

Differences in effects across AI types were significant (Q-between = 43.851; df = 3; $p < 0.001$). Mixed AI (k = 5) obtained the largest effect size, g = 1.640, as reflected in the studies of Eltahir & Babiker (2024), Zheng et al. (2024), Zablotska et al. (2026), Ayanwale & Omeh (2026), and Tan et al. (2025). The other category (k = 7) yielded g = 1.473, AI Chatbot (k = 11) yielded g = 0.986, and Generative AI (k = 5) yielded g = 0.865. Generative AI had the lowest internal heterogeneity (I^2 = 67.68%). The Adaptive Learning subgroup (k = 1) with the single study of Liu & Tu (2024) was not analyzed as it does not meet the requirement of $k \geq 2$.

DISCUSSION

The main finding of this meta-analysis is a large and significant effect size (g = 1.1971), confirming that AI makes a substantial contribution to the development of learners' critical thinking skills. This magnitude exceeds the earlier finding of Zhao et al., (2025), who reported g = 0.609 for the impact of Gen-AI on higher-order thinking in general, as well as syntheses showing moderate effects on higher-order thinking constructs. The difference in magnitude may reflect a combination of factors: (i) this study's exclusive focus on critical thinking as a more specific outcome construct than composite higher-order thinking (HOT); (ii) the selection of studies from a more contemporary publication period (2022–2026), reflecting the maturity of generative AI technology after ChatGPT; and (iii) the predominance of studies from higher education contexts in Asia, which generally apply AI within structured instructional designs.

This result is consistent with Vygotsky (1978) sociocultural perspective, which positions AI as adaptive scaffolding within the Zone of Proximal Development. As a cognitive partner, AI provides counter-argument questions, counterexamples, and requests for justification that encourage learners to engage in metacognitive reflection. (Hou et al., 2025) showed that dialogic interaction with AI prompts university students to re-verify arguments and evaluate evidence, which is the core of critical thinking. The phenomenon of AI hallucination, rather than being merely a weakness, can serve as a trigger for critical evaluation when it is integrated into an appropriate pedagogical design (Lee & Low, 2024).

However, the very high heterogeneity (I^2 = 95.08%) confirms that the effect of AI is not uniform but depends strongly on the implementation context. This is consistent with the arguments of Brazão & Tinoca, (2025) and Hassen, (2025) that the effectiveness of AI is not an inherent property of the technology, but a function of the quality of instructional design, the role of the facilitator, and learner readiness. The practical implication is that AI integration cannot be carried out generically; instead, interventions need to be designed with pedagogical contextuality in mind.

Moderator analyses help explain this variation, beginning with region. The largest effect size, found in studies from Africa (g = 1.887), is a finding that needs to be read with caution. Kooli et al. (2026), conducted in a higher education institution, reported g = 1.089 with the implementation of an AI chatbot accompanied by teacher training. Ayanwale & Omeh (2026), with g = 2.780, used a mixed AI approach combined with project-based learning. The intensive mentoring that accompanied the introduction of AI in Africa is likely a factor explaining the high magnitude. For Asia, the predominance of studies from higher education contexts in China and several other countries produced a significant and stable g = 1.121. The non-significant result for Europe (g = 0.672) should not be interpreted as ineffectiveness of AI; the fact that there are only two studies, and the strong contrast between Zablotska et al., (2026) and Bueno-Picazo & Tirado-Olivares (2026), indicates that the data are still too limited to draw conclusions.

Similar contextual differences appear across education levels. The higher education subgroup showed a significant and the most consistent effect size (g = 1.305). Park & Kim, (2025) found an improvement in critical thinking among students who interacted with an AI chatbot for argumentative discussion (g = 1.542). Luan et al., (2026) reported g = 1.260 in the context of developing university students' scientific reasoning, while Li & Long, (2025) obtained g = 1.079 using Generative AI for literature analysis tasks. At the senior high school level, although the mean effect size was fairly large (g = 1.068), the variation was also very high. . Aladini et al. (2025) recorded g = 0.939, Heydarnejad (2025) recorded g = 2.898, while Mai, (2026 recorded g = −1.840. This extreme contrast explains why the senior high school subgroup was not statistically significant. For elementary school, studies such as Dai et al. (2024) and Bueno-Picazo & Tirado-Olivares (2026) produced lower effect sizes, which is in line with the argument of Z. Liu et al. (2024) that elementary students require a more intensive level of scaffolding to optimize their interaction with AI.

Sustained scaffolding also depends on how long the intervention lasts. Interventions longer than 8 weeks produced g = 1.426, short durations of ≤4 weeks produced g = 1.118, and medium durations of 5–8 weeks produced the lowest g = 0.495. This asymmetric U-shaped pattern differs from the inverted-U curve reported by Zhao et al. (2025), which placed the peak effect at 8–16 weeks. Several explanations are possible. Short studies such as Baroud et al. (2025) (g = 2.678) and Nusivera et al. (2025) (g = 0.877) tend to benefit from the novelty effect of the technology. Medium-duration studies such as Li & Long, (2025)) (g = 1.079) and Cao et al. (2025) (g = 0.763) may be caught in a transition phase in which students are no longer very enthusiastic but have not yet fully integrated AI as a thinking tool. Long studies such as Wiboolyasarin et al. (2024) (g = 1.316) and Cao & Gao, (2026) (g = 2.786) allow enough time for the formation of reflective habits that accompany the use of AI. The practical implication is that program designers should consider a minimum duration of 8 weeks if they wish to establish stable critical thinking patterns.

Beyond duration, the type of AI also matters. Mixed AI, with the largest effect size (g = 1.640), signals that a multi-modal approach is more effective than a single intervention. Eltahir & Babiker, (2024) combined a chatbot, an adaptive system, and a generative module, while Zheng et al., (2024) integrated several AI tools to support reflective discussion. AI Chatbot, the most frequently studied category (k = 11), obtained a significant g = 0.986 but with high heterogeneity (I^2 = 96.63%). This large variation can be seen in the contrast between Essel et al. (2024) (g = 1.196), who used a ChatGPT chatbot with explicit instructional structure, and Mai, (2026) (g = −1.840), who used a chatbot without additional scaffolding. Generative AI had the lowest effect size (g = 0.865) but was the most consistent (I^2 = 67.68%). The studies by Li & Long (2025), Cai et al. (2026), Song et al. (2025), Luan et al. (2026), and Aladini et al. (2025) that used Generative AI produced relatively homogeneous effects, which may be attributed to the more uniform nature of Generative AI output, which students can evaluate more easily.

The overall results offer three practical implications. First, AI integration needs to be designed as a sustained program lasting at least eight weeks. The studies by Wiboolyasarin et al. (2024) and Heydarnejad (2025), which applied long interventions, clearly produced stronger effects. Second, combining several types of AI is recommended over using a single type, as seen in the large effects of the studies by Eltahir & Babiker (2024) and Zheng et al. (2024). Third, educators need to be equipped with adequate AI literacy and instructional design competence. The studies by Bueno-Picazo & Tirado-Olivares (2026) and Mai, (2026), which reported negative effects, serve as a warning that AI without pedagogical guidance can harm the development of students' critical thinking.

Despite these implications, several limitations should be noted. The predominance of studies from Asia (79.3%) limits the generalizability of the results to other contexts. The high heterogeneity ($I^2 = 95.08\%$) implies that the four moderators analyzed explain only part of the variation in effects; Future research could: (1) extend the coverage to Europe and the Americas to validate cross-regional patterns, (2) test the quality of AI instructional design as a moderator, and (4) distinguish the dimensions of critical thinking (analysis, evaluation, inference) to see which aspects are most affected by AI.

CONCLUSION

This meta-analysis of 29 quasi-experimental studies indicates that AI-based interventions are effective in improving students' critical thinking skills, with a large and statistically significant overall effect ($g = 1.1971$). No evidence of publication bias was detected, although this does not completely rule out its possibility. The size of the effect differed across region, educational level, treatment duration, and type of AI. A significant effect was found in higher education, in interventions longer than eight weeks, and in all AI types, with the largest estimate for combinations of several AI types. The larger effect in Africa rests on only four studies and should be interpreted cautiously. The very high heterogeneity suggests that effectiveness varies with the context in which AI is applied, although the four moderators explain only part of this variation. Several limitations apply: Asian studies dominate the sample, only quasi-experimental designs were included, and each moderator was analyzed separately. Future research should include studies from Europe and the Americas, test the quality of instructional design as a moderator, use longitudinal designs to examine long-term effects, and distinguish the dimensions of critical thinking.

REFERENCES

Abbas, R. M., Nser, H., & Farid, S. (2025). Exploring the Influence of AI-Generated Content on Critical Thinking Skills of University Students. Policy Research Journal, 3(5), 690–698. https://doi.org/10.5281/zenodo.15550335

Aladini, A., Ismail, S. M., Ahmad Saleem Khasawneh, M., & Shakibaei, G. (2025). Self-directed writing development across computer/AI-based tasks: Unraveling the traces on L2 writing outcomes, growth mindfulness, and grammatical knowledge. Computers in Human Behavior Reports, 17(December 2024), 100566. https://doi.org/10.1016/j.chbr.2024.100566

Alghamdi, L. H., & Alghizzi, T. M. (2025). Educators' reflections on AI-automated feedback in higher education: a structured integrative review of potentials, pitfalls, and ethical dimensions. Frontiers in Education, 10(November), 1–17. https://doi.org/10.3389/feduc.2025.1704820

Andrian, R. D., Widianti, Y. M., & Wati, D. A. R. (2025). The Role of Artificial Intelligence in Enhancing Critical Thinking Skills of Educational Technology Students: The Moderating Influence of Digital Literacy and Usage Regulation. Journal on Smart Learning Technologies, 1(2), 109–120.

Arimbawa, G. P. A., Parwati, N. N., Agustini, K., Suartama, I. K., & Sudarma, I. K. (2025).

Effectiveness Of Microlearning-Based E-Modules In Improving Critical Thinking Abilities Of Vocational High School Students. JKTP: Jurnal Kajian Teknologi Pendidikan, 8(3), 235–247. https://doi.org/10.17977/um038v8i32025p235-247

Ayanwale, M. A., & Omeh, C. B. (2026). AI-supported problem-based learning for enhancing computational thinking skills in STEM education. Computers in Human Behavior: Artificial Humans, 7(September 2025), 100263. https://doi.org/10.1016/j.chbah.2026.100263

Baroud, S., Alaoui, M. El, Idrissi, R. J., Boucetta, N., Ghariz, G., & Seghir, H. (2025). Improving biology learning in high school with ChatGPT: A case study in Morocco. Eurasia Journal of Mathematics, Science and Technology Education, 21(7). https://doi.org/10.29333/ejmste/16606

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. (2009a). Introduction to Meta-Analysis. John Wiley & Sons.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009b). Introduction to Meta-Analysis. In A Guide to the Scientific Career: Virtues, Communication, Research, and Academic Writing. https://doi.org/10.1002/9781118907283.ch68

Brazão, P., & Tinoca, L. (2025). Artificial intelligence and critical thinking: a case study with educational chatbots. Frontiers in Education, 10(September), 1–12. https://doi.org/10.3389/feduc.2025.1630493

Bueno-Picazo, M., & Tirado-Olivares, S. (2026). Generative AI in Geography Education: Content Creation and Conversational AI-Supported Learning to Promote Environmental Awareness. European Journal of Geography, 17(2), S.34-S.48. https://doi.org/10.48088/ejg.m.bue.17.2.034.048

Cai, J., Tan, Y. H., & Lin, S. E. (2026). Rethinking literature education: A quasi-experimental study of the production-oriented approach for english majors in China. Acta Psychologica, 266(March), 106808. https://doi.org/10.1016/j.actpsy.2026.106808

Cao, Y., & Gao, Q. (2026). Pedagogically Customized Machine Translation for Cultivating Critical Thinking in Foreign Language Education. Journal of Cases on Information Technology, 28(1), 1–17. https://doi.org/10.4018/JCIT.401921

Cohen, J. (1992). A power primer. Psychological Bulletin, 112(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

Dai, Y., Lin, Z., Liu, A., Dai, D., & Wang, W. (2024). Effect of an Analogy-Based Approach of Artificial Intelligence Pedagogy in Upper Primary Schools. Journal of Educational Computing Research, 61(8), 159–186. https://doi.org/10.1177/07356331231201342

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. BMJ, 315(7109), 629–634. https://doi.org/10.1136/bmj.315.7109.629

Eltahir, M. E., & Babiker, F. M. E. (2024). The Influence of Artificial Intelligence Tools on Student Performance in e-Learning Environments: Case Study. Electronic Journal of E-Learning, 22(9), 91–110. https://doi.org/10.34190/ejel.22.9.3639

Essel, H. B., Vlachopoulos, D., Essuman, A. B., & Amankwa, J. O. (2024). ChatGPT effects on cognitive skills of undergraduate students: Receiving instant responses from AI-based conversational large language models (LLMs). Computers and Education: Artificial Intelligence, 6(December 2023), 100198. https://doi.org/10.1016/j.caeai.2023.100198

Facione, P. A. (2011). Critical thinking: what it is and why it counts. Insight Assessment, 5(1), 1–23. https://www.researchgate.net/profile/Peter_Facione/publication/251303244_Critical_Thinking_What_It_Is_and_Why_It_Counts/links/5849b49608aed5252bcbe531/Critical-Thinking-What-It-Is-and-Why-It-Counts.pdf

Gao, J., Zhang, J., & Li, Y. (2025). Do AI chatbot-integrated writing tasks influence writing self-efficacy and critical thinking ability? An exploratory study. Computers and Education: Artificial Intelligence, 9(September 2024), 100472. https://doi.org/10.1016/j.caeai.2025.100472

García-Carmona, A. (2025). Scientific Thinking and Critical Thinking in Science Education : Two Distinct but Symbiotically Related Intellectual Processes. Science and Education, 34(1), 227–245. https://doi.org/10.1007/s11191-023-00460-5

Gerlich, M. (2025). AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking. Societies, 15(1), 1–28. https://doi.org/10.3390/soc15010006

Glass, G. V. (1976). Primary, Secondary, and Meta-Analysis of Research.

Hassen, M. (2025). The Impact of AI on Students' Reading, Critical Thinking, and Problem-Solving Skills. American Journal of Education and Information Technology, 9(2), 82–90. https://doi.org/10.11648/j.ajeit.20250902.12

Hedges, L. V., & Olkin, I. (1985). Statistical Methods for Meta-Analysis. Academic Press.

Heydarnejad, T. (2025). Unmasking the impacts of self-evaluation in AI-supported writing instruction on EFL learners' emotion regulation, self-competence, motivation, and writing achievement. Computers and Education: Artificial Intelligence, 9(June), 100494. https://doi.org/10.1016/j.caeai.2025.100494

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in knowledgebases. Journal of Intelligent Information Systems, 27(2), 159–184. https://doi.org/10.1007/s10844-006-2974-4

Hou, C., Zhu, G., & Sudarshan, V. (2025). The role of critical thinking on undergraduates' reliance behaviours on generative AI in problem-solving. British Journal of Educational Technology, 56(5), 1919–1941. https://doi.org/10.1111/bjet.13613

İdil, Ş., Gülen, S., & Dönmez, İ. (2024). What Should We Understand from PISA 2022 Results? Journal of STEAM Education, 7(1), 1–9. https://doi.org/10.55290/steam.1415261

Ilgun Dibek, M., Sahin Kursad, M., & Erdogan, T. (2025). Influence of artificial intelligence tools on higher order thinking skills: a meta-analysis. Interactive Learning Environments, 33(3), 2216–2238. https://doi.org/10.1080/10494820.2024.2402028

Jose, B., Cherian, J., Verghis, A. M., Varghise, S. M., S, M., & Joseph, S. (2025). The cognitive paradox of AI in education: between enhancement and erosion. Frontiers in Psychology, 16. https://doi.org/10.3389/fpsyg.2025.1550621

Kamel, H. (2024). Understanding the impact of AI Hallucinations on the university community. Cybrarians Journal, 73, 111–134. https://doi.org/10.70000/cj.2024.73.622

Khan, I., Ahmad, A. R., Jabeur, N., & Mahdi, M. N. (2021). An artificial intelligence approach to monitor student performance and devise preventive measures. Smart Learning Environments, 8(1). https://doi.org/10.1186/s40561-021-00161-y

Kooli, C., Yusuf, N., & Sarhan, M. Y. (2026). Colloquial engagement theory with AI awareness (CET-AIA): A new creative pedagogical framework for ethical assessment in the age of artificial intelligence. Thinking Skills and Creativity, 60(November 2025), 102051. https://doi.org/10.1016/j.tsc.2025.102051

Lee, C. C., & Low, M. Y. H. (2024). Using genAI in education: the case for critical thinking. Frontiers in Artificial Intelligence, 7. https://doi.org/10.3389/frai.2024.1452131

Li, C., & Long, J. (2025). Comparing AI-Assisted and Traditional Teaching in College English: Pedagogical Benefits and Learning Behaviors. Information (Switzerland), 16(10), 1–19. https://doi.org/10.3390/info16100895

Li, T., Krajcik, J. S., & Spiro, R. (2025). The Transformative Collaboration of Human Intelligence

and Artificial Intelligence in Designing Knowledge-in-Use Science Assessment for Learning. Journal of Science Education and Technology. https://doi.org/10.1007/s10956-025-10275-4

Li, Y., Kong, H., & Sciences, S. (2025). Education Studies Addressing " Hallucinations " in AI - Generated Content : Strategies for Developing Student Fact - Checking and. 1(2), 48–62.

Linsih, R. A., Soepriyanto, Y., & Praherdhiono, H. (2025). The Effect Of Collaborative Inquiry Based Learning Model And Learning Motivation On Critical Thinking Skills Of Elementary School Students. JKTP: Jurnal Kajian Teknologi Pendidikan, 8(2), 118–127. https://doi.org/10.17977/um038v8i22025p118

Liu, Q., & Tu, C. C. (2024). Improving Critical Thinking through AI-Supported Socio-Scientific Issues Instruction. Journal of Logistics, Informatics and Service Science, 11(3), 52–65. https://doi.org/10.33168/JLISS.2024.0304

Liu, Z., Gao, Y., Yang, Y., Kong, X., & Zhao, L. (2024). Exploring the influence of students' perceptions of face-to-face collaboration on cognitive engagement and learning outcomes in collaborative programming. Acta Psychologica, 248(July), 104393. https://doi.org/10.1016/j.actpsy.2024.104393

Luan, J., Halim, H. A., & Saad, M. R. M. (2026). The Effect of AI and Multiple Intelligences in Project-Based Learning on English Achievement. International Journal of Computer-Assisted Language Learning and Teaching, 16(1), 1–28. https://doi.org/10.4018/IJCALLT.400912

Mai, H. T. (2026). Evaluating the impact of using ChatGPT on EFL students' argumentative skills. Discover Education, 5(1), 1–20. https://doi.org/10.1007/s44217-026-01153-1

Maia, J. de S. Z., Bueno, A. P. A., & Sato, J. R. (2023). Applications of Artificial Intelligence Models in Educational Analytics and Decision Making: A Systematic Review. World, 4(2), 288–313. https://doi.org/10.3390/world4020019

Najah, E. F., Syaharuddin, Gatan, P. R., Mohammed, R. A., & Abdillah. (2024). The role of Information Literacy in Enhancing Students' Critical Thinking Skills. Jurnal Literasi Perpustakaan Dan Informasi UHO: Jurnal Penelitian Kajian Perpustakaan Dan Informasi, 4(3), 129–139. https://doi.org/10.52423/jlpi.v4i3.15

Nusivera, E., Hikmat, A., & Ghani, A. R. A. (2025). Integration of Chat-GPT Usage in Language Learning Model to Improve Argumentation Skills, Complex Comprehension Skills, and Critical Thinking Skills. International Journal of Learning, Teaching and Educational Research, 24(2), 375–390. https://doi.org/10.26803/ijlter.24.2.19

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. Bmj, 372. https://doi.org/10.1136/bmj.n71

Park, S. A., & Kim, H. Y. (2025). Development and effects of a scenario-based labor nursing simulation education program using an artificial intelligence tutor: a quasi-experimental study. Women's Health Nursing, 31(2), 143–154. https://doi.org/10.4069/whn.2025.06.18

Partnership for 21 st Century Skills. (2015). Partnership for 21St Century Skills-Core Content Integration. Ohio Department of Education, 1. www.P21.org.

Piaget, J. (1970). Science of Education and the Psychology of the Child. Orion Press. https://books.google.co.id/books?id=ZwedAAAAMAAJ

Sinaga, A. I., Nasution, A. F., & Albina, M. (2025). Effect Of Artificial Intelligence Use On Creativity In Writing Scientific Works. JKTP: Jurnal Kajian Teknologi Pendidikan, 8(3), 226–234. https://doi.org/10.17977/um038v8i32025p226-234

Song, D., Zhang, P., Zhu, Y., Qi, S., Yang, Y., Gong, L., & Zhou, L. (2025). Effects of generative artificial intelligence on higher-order thinking skills and artificial intelligence literacy in

nursing undergraduates: A quasi-experimental study. Nurse Education in Practice, 88(81), 104549. https://doi.org/10.1016/j.nepr.2025.104549

Sun, X., Liu, C., Li, P., & Cai, Z. (2026). Effects of Generative AI Tools On Students' Critical Thinking: A Three-Level Meta-Analysis. Educational Psychology Review, 38(1), 84. https://doi.org/10.1007/s10648-026-10180-1

Sun, Y., Sheng, D., Zhou, Z., & Wu, Y. (2024). AI hallucination: towards a comprehensive classification of distorted information in artificial intelligence-generated content. Humanities and Social Sciences Communications, 11(1), 1–14. https://doi.org/10.1057/s41599-024-03811-x

Tan, X., Cheng, G., & Ling, M. H. (2025). Enhancing teachers' AI competency: A professional development intervention study based on intelligent-TPACK framework. Computers and Education: Artificial Intelligence, 9(November), 100521. https://doi.org/10.1016/j.caeai.2025.100521

Tian, J., & Zhang, R. (2025). Learners' AI dependence and critical thinking: The psychological mechanism of fatigue and the social buffering role of AI literacy. Acta Psychologica, 260(October), 105725. https://doi.org/10.1016/j.actpsy.2025.105725

Vygotsky, L. S. (1978). Mind in Society: Development of Higher Psychological Processes. Harvard University Press. https://books.google.co.id/books?id=RxjjUefze_oC

Wiboolyasarin, W., Wiboolyasarin, K., Suwanwihok, K., Jinowat, N., & Muenjanchoey, R. (2024). Synergizing collaborative writing and AI feedback: An investigation into enhancing L2 writing proficiency in wiki-based environments. Computers and Education: Artificial Intelligence, 6(March), 100228. https://doi.org/10.1016/j.caeai.2024.100228

Yas, N., Abdulrahim, H., Rabadi, I. G., & Abdijabar, Z. (2025). Protecting the civil rights of the author from the risks of artificial intelligence. Research Journal in Advanced Humanities, 6(1), 1–9. https://doi.org/10.58256/s7rhdm72

Zablotska, L., Chernii, L., Meleshchenko, V., Olendr, T., & Tsar, I. (2026). Developing Students' Critical Thinking through AI-Enhanced Foreign Language Instruction. Arab World English Journal, 3, 132–146. https://doi.org/10.24093/awej/ai3.9

Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: a systematic review. Smart Learning Environments, 11(1). https://doi.org/10.1186/s40561-024-00316-7

Zhao, Y., Yue, Y., Sun, Z., Jiang, Q., & Li, G. (2025). Does Generative Artificial Intelligence Improve Students' Higher-Order Thinking? A Meta-Analysis Based on 29 Experiments and Quasi-Experiments. Journal of Intelligence, 13(12), 1–21. https://doi.org/10.3390/jintelligence13120160

Zheng, K., Shen, Z., Chen, Z., Che, C., & Zhu, H. (2024). Application of AI-empowered scenario-based simulation teaching mode in cardiovascular disease education. BMC Medical Education, 24(1). https://doi.org/10.1186/s12909-024-05977-z