International Education Trend Issues
Volume 3, Number 1, 2025
https://ijble.com/index.php/ieti

Education at the Cultural Crossroads: Navigating Identity and Globalization in Southeast Asia

Yosphia Fahruddiana
Universitas Sultan Ageng Tirtayasa
Correspondence: Yosphia@gmail.com

Abstract
Artificial intelligence (AI)-assisted learning is expanding rapidly, but achievement effects vary across tools, outcomes, and implementation settings. This study systematically audited a supplied Scopus corpus and conducted an updated second-order random-effects meta-analysis. The export contained 200 publications from 2024. Nineteen records met a sensitive AI/education concept screen, 11 were assessed in detail, and six instructional or learning-support studies were retained for narrative synthesis. Because the Scopus metadata did not provide complete condition-level statistics for standardized effect calculation, no primary effect was invented. The quantitative component therefore pooled six verified first-order meta-analyses using restricted maximum likelihood and Hartung-Knapp inference. The overall effect was positive and moderate (Hedges' g = 0.553, 95% CI [0.392, 0.713]); heterogeneity was moderate (I^2 = 52.9%, τ^2 = 0.009), and the 95% prediction interval remained positive [0.265, 0.840]. Leave-one-out estimates ranged from 0.505 to 0.585. Benefits were most plausible when AI supported adaptive practice, dialogue, feedback, and reflection within teacher-guided, assessment-aligned designs. Primary-study overlap and incomplete metadata limit causal generalization.

Keywords: artificial intelligence in education; student achievement; meta-analysis; generative AI; intelligent tutoring systems

1. INTRODUCTION

Artificial intelligence in education has moved from narrowly defined tutoring programs toward a heterogeneous ecology of adaptive courseware, conversational agents, automated feedback systems, learning analytics, and generative AI. These technologies differ substantially in what they automate and in how closely their decisions are integrated with instruction. Consequently, the educational question is not whether an algorithm is present, but whether the system changes the quality, timing, or personalization of learning activity (Ouyang & Jiao, 2021; Roll & Wylie, 2016).

Achievement effects are theoretically plausible. AI systems can diagnose errors, select tasks at an appropriate level, provide immediate explanation, and increase opportunities for deliberate practice. Intelligent tutoring systems have long shown benefits over conventional instruction, although the magnitude of improvement depends on comparison conditions, instructional domain, and study quality (Kulik & Fletcher, 2016; Ma et al., 2014). More recent chatbots and generative systems can extend these mechanisms through natural-language dialogue, worked examples, and iterative feedback, but they can also provide inaccurate information, encourage superficial completion, or weaken independent reasoning when learners use them without epistemic monitoring (Kasneci et al., 2023; Tlili et al., 2023).

The empirical literature has expanded faster than a stable consensus has formed. Reviews have documented rapid growth in higher education, yet they also identify limited educator involvement, inconsistent ethical reporting, and a concentration of research in well-resourced settings (Bond et al., 2024; Crompton & Burke, 2023; Zawacki-Richter et al., 2019). First-order meta-analyses generally report positive achievement effects, but their scopes range from intelligent tutoring systems and AI-enabled STEM personalization to blended learning, ChatGPT, and generative AI. Pooling such evidence requires explicit attention to construct boundaries and overlap among the primary studies represented in the reviews.

The supplied Scopus export illustrates a related evidence problem. Its 200 records contain controlled LLM use, pedagogical agents, automated feedback, learning analytics, AI-supported writing and mathematics, recommender-system evaluation, responsible-AI reviews, and many records in which AI is only a background term. Predicting achievement, measuring model accuracy, or teaching AI as course content is educationally relevant, but none alone demonstrates that AI-assisted instruction raises achievement. The instructional studies in the corpus therefore required separate appraisal (Arguedas et al., 2024; Canonigo, 2024; Kumar et al., 2024; Teng, 2024).

This study addresses that distinction through a transparent two-part design. First, it audits the supplied Scopus corpus and synthesizes the eligible learning-support studies narratively. Second, because the export does not contain the numerical fields required for a defensible first-order effect calculation, it conducts an updated second-order meta-analysis of verified first-order meta-analyses. The study asks: (1) Which records in the supplied export represent substantive AI-assisted learning? (2) What opportunities and limitations emerge from the eligible primary studies? (3) What is the pooled synthesis-level effect of AI-assisted learning on student achievement? (4) How stable is the estimate under leave-one-out analysis, and what implementation conditions plausibly explain variation?

3. METHOD

The study used a PRISMA-informed systematic audit followed by a second-order random-effects meta-analysis. PRISMA principles guided transparent reporting of the evidence flow, eligibility decisions, and reasons for exclusion (Page et al., 2021). The review was not prospectively registered; this is acknowledged as a limitation. The attached CSV file was treated as the user-supplied discovery corpus, whereas the quantitative model used verified effect estimates from first-order meta-analyses.

The supplied Scopus export contained 200 records, all published in 2024. Titles, abstracts, author keywords, index keywords, source titles, and document types were screened in two stages: a sensitive concept screen followed by substantive assessment of whether AI directly supported learning. The original Scopus query string and search-history log were not embedded in the export; consequently, the file is reported as a bounded evidence corpus rather than as proof of exhaustive database coverage. Publisher pages and bibliographic discovery services were additionally checked through 2 August 2026 to verify first-order meta-analyses that reported an overall standardized effect and a confidence interval for AI-assisted learning outcomes.

Table 1. Eligibility criteria for the quantitative synthesis

Dimension	Included	Excluded
Population	Students or learners in formal or structured educational settings	Non-learning populations; purely clinical or organizational samples
Intervention	AI system with an instructional function: tutoring, adaptation, dialogue, feedback, or guided generation	Prediction, detection, administration, or content moderation without an instructional treatment
Comparator/ outcome	Achievement, knowledge, skills, grades, or cognitive learning outcome with a comparator	Attitudes, intention, perceptions, or prediction accuracy alone
Design	First-order meta-analysis reporting an overall standardized effect and uncertainty interval	Narrative review, bibliometric study, protocol, or single primary study
Reporting	Effect estimate convertible to Hedges-type g with a 95% CI	Insufficient numerical reporting or duplicate synthesis

Two conceptual screens were applied. The first identified records that explicitly concerned AI, machine learning, large language models, neural networks, or algorithmic educational systems. The second determined whether AI was the instructional treatment and whether controlled achievement statistics were extractable. Records on academic-performance prediction, source searching, content detection, or perceptions were retained in the audit but excluded from the meta-analysis. This separation prevents predictive-model performance from being misinterpreted as a learning effect.

For each eligible first-order meta-analysis, the following were extracted: scope, number of studies or independent units, pooled effect, lower and upper 95% confidence limits, and publication details. Estimates were treated as Hedges-type standardized mean differences. Potential overlap among primary studies was assessed conceptually from the reported scopes and years; because complete study-level overlap matrices were not consistently available, a formal corrected covered area could not be calculated.

The standard error for each synthesis-level estimate was derived as SE = (upper CI - lower CI)/(2 × 1.96), and sampling variance was SE^2. A random-effects model was fitted because the meta-analyses addressed different AI systems, educational levels, and outcomes. Between-synthesis variance (τ^2) was estimated by restricted maximum likelihood. The pooled confidence interval used the Hartung-Knapp adjustment, which is preferable to a normal approximation when the number of estimates is small (Borenstein et al., 2021; IntHout et al., 2014).

Heterogeneity was summarized using Cochran's Q, I^2, and τ^2 (Higgins & Thompson, 2002). A 95% prediction interval estimated the range in which the underlying effect of a comparable future synthesis might fall. Leave-one-out analysis assessed whether any single meta-analysis dominated the pooled result. Publication-bias tests were not performed because six synthesis-level estimates provide insufficient power and violate the independence assumptions of conventional funnel-plot methods. Computation followed standard meta-analytic formulas (Hedges & Olkin, 1985; Viechtbauer, 2010).

4. RESULTS

4.1 Audit of the supplied Scopus corpus

The export contained 200 records from 2024. Nineteen records met the sensitive AI/education concept screen. Eleven were assessed in detail as substantive AI or learning-analytics applications, and six learning-support studies were retained for narrative synthesis. The remaining concept hits concerned non-AI educational technology, AI as curriculum content, or incidental terminology. None of the six narrative studies supplied complete condition-level means, standard deviations, and sample sizes in the Scopus metadata; therefore, no attached record was converted into a standardized effect size.

Table 2. Classification of substantive AI-education records in the supplied corpus

Audit category	n	Interpretive status
AI-assisted instructional intervention	4	Direct instructional assistance; retained narratively, but metadata were insufficient for primary effect extraction.
Analytics-assisted learning intervention	2	Learning analytics formed part of the intervention; retained narratively with outcome-boundary cautions.
Automated personalized feedback	1	Randomized feedback study with motivational and emotional outcomes rather than achievement.
AI recommender-system evaluation	1	System accuracy was evaluated; not an achievement treatment effect.
AI-supported educational citizen science	1	AI supported the project, but the quantitative outcome was model classification accuracy.
Responsible-AI systematic review	1	Contextual evidence on ethics and human-centred implementation, not a primary effect.
Learning-analytics observational study	1	Behavioural associations were analyzed without an AI instructional contrast.

The narrative evidence was directionally positive but heterogeneous. Guided LLM use produced only marginal performance gains and showed that interaction design changed trust and query behaviour (Kumar et al., 2024). An affective pedagogical tutor improved several cognitive-feedback ratings relative to human feedback, although the dependent measure largely reflected students' perceptions of learning effectiveness (Arguedas et al., 2024). Analytics-assisted reflective assessment and visual learning analytics supported epistemic agency, productive inquiry, and domain knowledge (Tong et al., 2024; Yang et al., 2024). ChatGPT-supported writing and AI-supported mathematics were associated with higher motivation, self-efficacy, conceptual understanding, or engagement, but their metadata did not permit controlled standardized-effect calculation (Canonigo, 2024; Teng, 2024).

The audit also identified a randomized study of automated personalized feedback in which reference-frame design altered motivation and achievement emotions rather than academic achievement (Weidlich et al., 2024). Recommender-system accuracy, machine-learning classification, responsible-AI reviews, and observational learning analytics were retained as contextual evidence but excluded from causal achievement synthesis (Fu & Weng, 2024).

4.2 Characteristics of the included meta-analyses

Six first-order meta-analyses met the quantitative criteria. Their scopes ranged from K-12 personalized STEM education and AI-enhanced blended learning to generative AI, ChatGPT, and broad classroom applications. The reported effects ranged from g = 0.449 to g = 0.924. This spread supported a random-effects model rather than a common-effect assumption.

Table 3. First-order meta-analyses included in the second-order synthesis

Source	Scope	Studies / units	g	95% CI
Liu et al. (2026)	Broad AI-assisted learning	49	0.449	[0.194, 0.704]
Li et al. (2025)	AI-enabled personalized K-12 STEM	99	0.455	[0.327, 0.583]
Wu et al. (2025)	AI-enhanced blended learning	21	0.500	[0.270, 0.740]
Fan et al. (2026)	Generative AI in higher education	36	0.499	[0.372, 0.627]
Wu et al. (2026)	ChatGPT-supported learning outcomes	35	0.670	[0.495, 0.844]
Dong et al. (2025)	AI and academic achievement	29	0.924	[0.612, 1.235]

The included values were taken from verified reports on broad AI-assisted learning, personalized STEM, blended learning, generative AI, ChatGPT, and classroom achievement (Dong et al., 2025; Fan et al., 2026; Li et al., 2025; Liu et al., 2026; J. Wu et al., 2025; X. Wu et al., 2026).

4.3 Overall effect and heterogeneity

The REML-Hartung-Knapp model produced a pooled effect of g = 0.553 (95% CI [0.392, 0.713]). The confidence interval excluded zero, indicating a positive synthesis-level association between AI-assisted learning and student achievement. Heterogeneity was moderate: $Q(5) = 10.61$, $p = 0.060$, $I^2 = 52.9\%$, and $\tau^2 = 0.009$. The 95% prediction interval [0.265, 0.840] remained positive, although its width indicates meaningful contextual variation.

Table 4. Summary of the second-order random-effects model

Statistic	Estimate
Number of meta-analyses	6
Pooled Hedges-type g	0.553
95% Hartung-Knapp CI	[0.392, 0.713]
Cochran Q (df = 5)	10.61, p = 0.060
I^2	52.9%
τ^2	0.009
95% prediction interval	[0.265, 0.840]

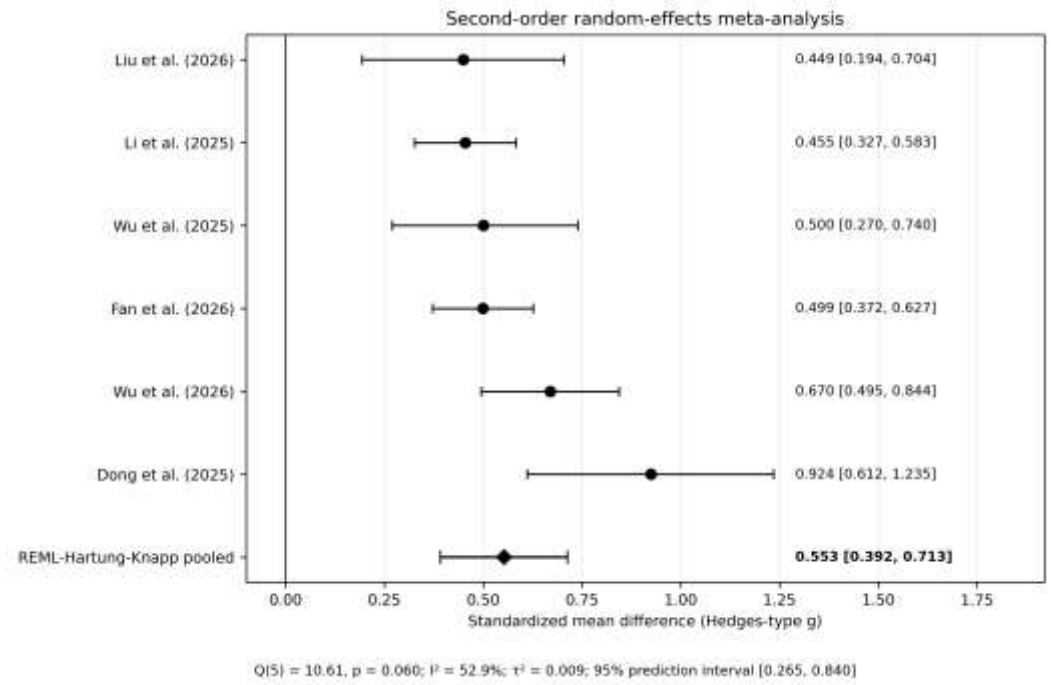

Figure 3. Forest plot of synthesis-level effects and the pooled estimate.

4.4 Sensitivity analysis

Leave-one-out estimates ranged from 0.505 to 0.585, and every confidence interval remained above zero. Omitting the broad classroom meta-analysis by Dong et al. reduced heterogeneity substantially, indicating that its comparatively large estimate contributed strongly to between-synthesis variation. Nevertheless, removing it did not change the substantive conclusion of a positive effect.

Table 5. Leave-one-out sensitivity results

Omitted synthesis	Re-estimated g	95% CI	I^2
Liu et al. (2026)	0.573	[0.367, 0.779]	60.7%
Li et al. (2025)	0.585	[0.376, 0.794]	53.5%
Wu et al. (2025)	0.568	[0.352, 0.785]	62.0%
Fan et al. (2026)	0.577	[0.351, 0.803]	60.9%
Wu et al. (2026)	0.505	[0.352, 0.657]	48.3%
Dong et al. (2025)	0.512	[0.405, 0.619]	4.6%

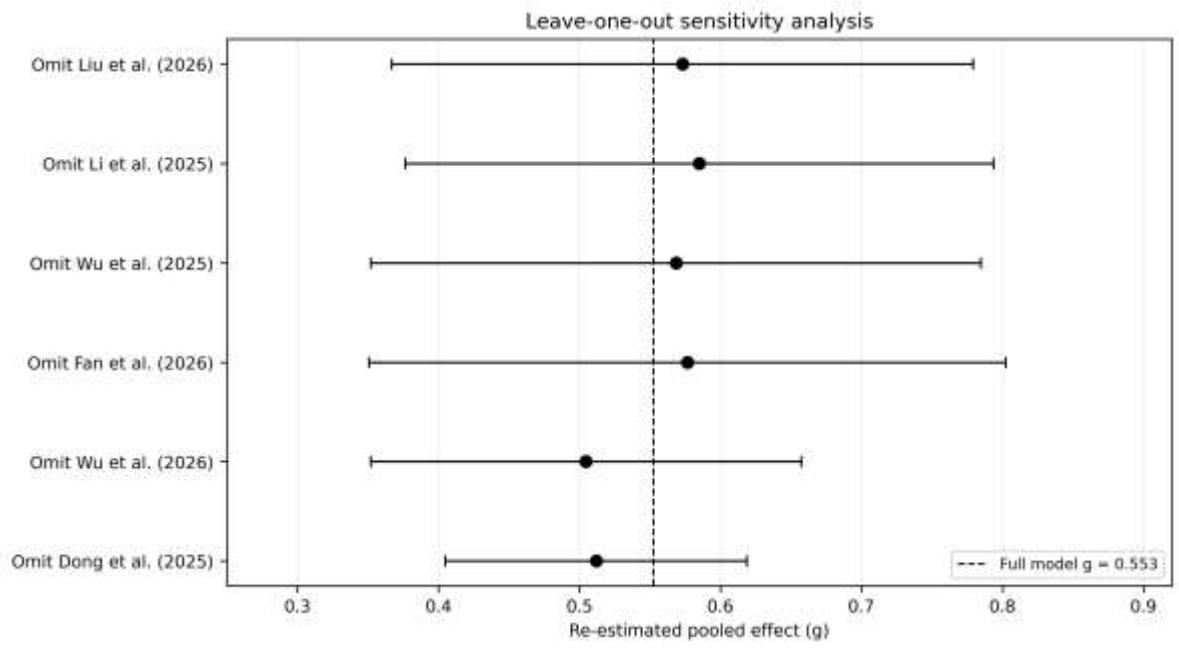

Figure 4. Leave-one-out re-estimation of the pooled effect.

5. DISCUSSION

5.1 Principal finding

The central finding is a moderate positive synthesis-level effect of AI-assisted learning on student achievement. The estimate is more conservative than the largest broad meta-analysis and higher than the smallest domain-specific estimates, which is expected when heterogeneous reviews are integrated. The positive prediction interval suggests that a comparable future synthesis is more likely to find benefit than harm, but the interval should not be interpreted as a guarantee for every tool, learner, or institution.

The result also reinforces a methodological distinction: evidence that an algorithm predicts achievement is not evidence that it improves achievement. The attached Scopus corpus contained more AI-adjacent, observational, system-evaluation, and contextual research than extractable controlled achievement evidence. Treating those records as if they supplied treatment effects would have produced a numerically impressive but conceptually invalid meta-analysis.

5.2 Why effects vary

Variation is consistent with differences in AI function and instructional design. Personalized systems can improve task selection and practice density; tutors can provide hints and worked steps; chatbots can support question generation and dialogue; and generative AI can provide rapid drafts, explanations, or feedback. However, these affordances operate through learning activity. Benefits are less likely when students copy answers, when feedback is inaccurate or too generic, or when assessment does not reward the targeted knowledge and skills (Kasneci et al., 2023; Tlili et al., 2023; UNESCO, 2023).

The included meta-analyses reported moderators involving technology type, educational level, subject, intervention duration, instructional mode, and measurement. Personalized systems in blended learning sometimes outperformed generic chatbots, while K-12 STEM estimates varied across school level and application type. The practical implication is not to select AI solely by brand or model capability. Institutions should specify the learning problem, determine the pedagogical function of AI, and align feedback and assessment with that function.

5.3 Implications for educational practice and policy

Table 6. Evidence-informed implementation implications

Implementation domain	Recommended practice	Rationale
Instructional alignment	Define the target knowledge or skill before selecting the AI function.	Effectiveness depends on whether tutoring, generation, or feedback serves the intended outcome.
Teacher orchestration	Use teacher-set prompts, checkpoints, and discussion of errors.	Human guidance converts AI output into monitored learning activity.
Assessment design	Require explanation, revision, source checking, and transfer tasks.	Assessment should reward learning rather than answer production.
Equity and access	Provide devices, connectivity, language support, and accessible alternatives.	Average effects can conceal unequal opportunity to benefit.
Data and ethics	Minimize personal data, disclose AI use, and establish verification routines.	Trustworthy adoption requires privacy, transparency, and error management.
Evaluation	Use pre-specified outcomes, appropriate comparators, and sufficient intervention duration.	Stronger designs improve causal inference and cross-study comparability.

For educators, the evidence supports cautious adoption rather than blanket enthusiasm or rejection. AI should be treated as an instructional component whose contribution is tested against a credible alternative. For institutions, procurement criteria should include curricular alignment, explainability, data governance, accessibility, and evidence from comparable learners. For policy makers, professional development and infrastructure are necessary to prevent benefits from accumulating only in already advantaged settings (Bond et al., 2024; UNESCO, 2023).

5.4 Limitations and research agenda

Several limitations constrain interpretation. First, the attached export was broad and lacked the original search string, so it cannot demonstrate comprehensive coverage of primary AI-learning interventions. Second, the quantitative unit was the meta-analysis, not the primary study. Primary-study overlap across reviews may narrow the effective evidence base and make conventional standard errors optimistic. Third, the six estimates combine different outcomes and AI categories; this is the rationale for random effects, but it also limits causal specificity. Fourth, with $k = 6$, moderator analysis and publication-bias tests would be underpowered. Fifth, the rapid pace of generative-AI research means that new controlled studies may alter the estimate.

Future reviews should preregister a multi-database search, archive full strategies, obtain primary-study lists from each meta-analysis, quantify overlap, and conduct robust variance or multilevel models when dependent effects are available. Primary studies should report group means, standard deviations, sample sizes, attrition, implementation fidelity, and AI interaction data. Longer follow-up is needed to distinguish novelty effects from durable learning and to test whether AI strengthens transfer, critical thinking, and independent performance after assistance is removed.

6. CONCLUSION

This systematic review found a moderate positive synthesis-level effect of AI-assisted learning on student achievement ($g = 0.553$), with a positive prediction interval and stable leave-one-out estimates. The result supports the educational potential of adaptive practice, tutoring, dialogue, and feedback, but it does not justify treating every AI application as effective. The supplied Scopus corpus showed that much AI-and-academic-performance research concerns prediction rather than intervention. A defensible evidence synthesis must preserve that distinction. AI is most likely to improve achievement when its function is pedagogically explicit, teacher-guided, assessment-aligned, accessible, and evaluated with transparent controlled designs.

REFERENCES

Arguedas, M., Daradoumis, T., & Caballé, S. (2024). Measuring the effects of pedagogical agent cognitive and affective feedback on students' academic performance. *Frontiers in Artificial Intelligence*, *7*, 1495342. https://doi.org/10.3389/frai.2024.1495342

Bond, M., Khosravi, H., de Laat, M., Bergdahl, N., Negrea, V., Oxley, E., Pham, P., Chong, S. W., & Siemens, G. (2024). A meta systematic review of artificial intelligence in higher education: A call for increased ethics, collaboration, and rigour. *International Journal of Educational Technology in Higher Education*, *21*(1), 4. https://doi.org/10.1186/s41239-023-00436-z

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2021). *Introduction to Meta-Analysis*. Wiley. https://doi.org/10.1002/9781119558378

Canonigo, A. M. (2024). Levering AI to enhance students' conceptual understanding and confidence in mathematics. *Journal of Computer Assisted Learning*, *40*(6), 3215–3229. https://doi.org/10.1111/jcal.13065

Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education*, *20*, 22. https://doi.org/10.1186/s41239-023-00392-8

Dong, L., Tang, X., & Wang, X. (2025). Examining the effect of artificial intelligence in relation to students' academic achievement: A meta-analysis. *Computers and Education: Artificial Intelligence*, *8*, 100400. https://doi.org/10.1016/j.caeai.2025.100400

Fan, C., Ke, L., Chen, Z., & Lv, P. (2026). Exploring the effect of GenAI on learning outcomes in higher education: A three-level meta-analysis. *Frontiers in Psychology*, *17*, 1758670. https://doi.org/10.3389/fpsyg.2026.1758670

Fu, Y., & Weng, Z. (2024). Navigating the ethical terrain of AI in education: A systematic review on framing responsible human-centered AI practices. *Computers and Education: Artificial Intelligence*, *7*, 100306. https://doi.org/10.1016/j.caeai.2024.100306

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

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, *21*(11), 1539–1558. https://doi.org/10.1002/sim.1186

IntHout, J., Ioannidis, J. P. A., & Borm, G. F. (2014). The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. *BMC Medical Research Methodology*, *14*, 25. https://doi.org/10.1186/1471-2288-14-25

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (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

Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of intelligent tutoring systems: A meta-analytic review. *Review of Educational Research*, *86*(1), 42–78. https://doi.org/10.3102/0034654315581420

Kumar, H., Musabirov, I., Reza, M., Shi, J., Wang, X., Williams, J. J., Kuzminykh, A., & Liut, M. (2024). Guiding Students in Using LLMs in Supported Learning Environments: Effects on Interaction Dynamics, Learner Performance, Confidence, and Trust. *Proceedings of the ACM on Human-Computer Interaction*, *8*(CSCW2), 499. https://doi.org/10.1145/3687038

Li, S., Zeng, C., Liu, H., Jia, J., Liang, M., Cha, Y., Lim, C. P., & Wu, X. (2025). A meta-analysis of AI-enabled personalized STEM education in schools. *International Journal of STEM Education*, *12*, 58. https://doi.org/10.1186/s40594-025-00566-y

Liu, C., Meng, H., Zhu, H., & Rong, C. (2026). Does AI-assisted learning improve students' learning outcomes? Evidence from a meta-analysis. *The Asia-Pacific Education Researcher*, *35*, 1021–1033. https://doi.org/10.1007/s40299-026-01104-2

Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of Educational Psychology*, *106*(4), 901–918. https://doi.org/10.1037/a0037123

Ouyang, F., & 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, 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*, n71. https://doi.org/10.1136/bmj.n71

Roll, I., & Wylie, R. (2016). Evolution and revolution in artificial intelligence in education. *International Journal of Artificial Intelligence in Education*, *26*(2), 582–599. https://doi.org/10.1007/s40593-016-0110-3

Teng, M. F. (2024). {ChatGPT is the companion, not enemies}: EFL learners' perceptions and experiences in using ChatGPT for feedback in writing. *Computers and Education: Artificial Intelligence*, *7*, 100270. https://doi.org/10.1016/j.caeai.2024.100270

Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*, *10*, 15. https://doi.org/10.1186/s40561-023-00237-x

Tong, Y., Yang, C., & Chen, G. (2024). A visual learning analytics approach for knowledge building: Impact on students' epistemic understanding of discourse, productive inquiry and domain knowledge. *British Journal of Educational Technology*, *55*(3), 992–1019. https://doi.org/10.1111/bjet.13409

UNESCO. (2023). *Guidance for Generative AI in Education and Research*. UNESCO.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03

Weidlich, J., Fink, A., Jivet, I., Yau, J., Giorgashvili, T., Drachsler, H., & Frey, A. (2024). Emotional and motivational effects of automated and personalized formative feedback: The role of reference frames. *Journal of Computer Assisted Learning*, *40*(6), 2735–2752. https://doi.org/10.1111/jcal.13024

Wu, J., Tlili, A., Salha, S., Mizza, D., Saqr, M., López-Pernas, S., & Huang, R. (2025). Unlocking the potential of artificial intelligence in improving learning achievement in blended learning: A meta-analysis. *Frontiers in Psychology*, *16*, 1691414. https://doi.org/10.3389/fpsyg.2025.1691414

Wu, R., & 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

Wu, X., Zhu, P., Zhang, J., Yin, M., & Wang, Y. (2026). ChatGPT's impact on student learning outcomes: A meta-analysis of 35 experimental studies. *Humanities and Social Sciences Communications*, *13*, 684. https://doi.org/10.1057/s41599-026-07019-z

Yang, Y., Feng, X., Zhu, G., & Xie, K. (2024). Effects and mechanisms of analytics-assisted reflective assessment in fostering undergraduates' collective epistemic agency in computer-supported collaborative inquiry. *Journal of Computer Assisted Learning*, *40*(3), 1098–1122. https://doi.org/10.1111/jcal.12915

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