Natural Resources for Human Health
eISSN: 2583-1194

Original Research
Received 06 May 2026
Revised 17 June 2026
Accepted 30 June 2026

Natural Resources for Human Health 2026, Vol. 6 (11s): 262–274

KEYWORDS:
Generative Artificial Intelligence, Health Science Education, Meta-Analysis, Critical Thinking, Problem-Solving, Clinical Reasoning, Decision-Making

Effects of Generative Artificial Intelligence on Higher-Order Cognitive Outcomes in Health Science Education: A Meta-Analysis of Interventional Studies

Anas Ali Alhur[1], Nouf Khalid Al-Kahtani[1], Emad Abdul-Mahsoud Mahgoub[2]

[1] Department of Health Information Management and Technology, College of Public Health, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia
[2] Department of Health Psychology, College of Public Health, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia
Corresponding author: Anas Ali Alhur
Email: 2250700045@iau.edu.sa, ORCID: 0000-0001-6044-7072

ABSTRACT:

Background: Generative artificial intelligence (GenAI) is increasingly used in health science education, but its effects on higher-order cognitive outcomes remain uncertain.

Objective: To quantify the effects of GenAI-supported educational interventions on critical thinking, problem-solving, clinical reasoning, and decision-making among health science learners.

Methods: This meta-analysis used quantitatively eligible interventional studies identified within a prospectively registered evidence-synthesis protocol (PROSPERO CRD420251128495). PubMed, Scopus, the Cochrane Library, and relevant gray literature were searched between September and November 2025. Studies were eligible for quantitative synthesis when they evaluated a GenAI-supported educational intervention and reported sufficient data to calculate standardized mean differences. Hedges' g with 95% confidence intervals (CIs) was synthesized using DerSimonian-Laird random-effects models. Heterogeneity was assessed using Cochran's Q, I^2, and τ^2. Publication bias and small-study effects were explored using funnel plots and Begg's and Egger's tests, and robustness was examined using leave-one-effect-estimate-out sensitivity analysis.

Results: Fifteen unique interventional studies contributed 21 outcome-domain effect estimates to the revised meta-analysis after exclusion of Wu et al., whose conventional AI diagnostic platform did not meet the operational definition of GenAI. The overall pooled effect favored GenAI-supported education (Hedges' g = 0.58, 95% CI 0.22–0.94, p = 0.002), with considerable heterogeneity (Q = 475.07, p < 0.001; I^2 = 95.79%; τ^2 = 0.651). Decision-making showed a large positive pooled effect (g = 1.52, 95% CI 0.78–2.26), and critical thinking also showed a large positive effect (g = 1.75, 95% CI 1.08–2.41). Problem-solving showed a small negative pooled effect (g = −0.22, 95% CI −0.32 to −0.12), whereas clinical reasoning was not statistically significant (g = −0.19, 95% CI −0.79 to 0.41). Begg's (p = 0.084) and Egger's (p = 0.176) tests were not statistically significant. Leave-one-effect-estimate-out analyses produced pooled estimates ranging from 0.47 to 0.65.

Natural Resources for Human Health
eISSN: 2583-1194

Conclusions: GenAI-supported education was associated with a moderate overall improvement in higher-order cognitive outcomes, but effects differed substantially across domains. The very high heterogeneity, together with dependence among multiple effect estimates contributed by some studies, requires cautious interpretation of the overall pooled estimate. Domain-specific findings suggest potential benefits for decision-making and critical thinking, while problem-solving and clinical reasoning require more careful instructional design and further study.

INTRODUCTION

Generative artificial intelligence (GenAI) technologies such as ChatGPT, Gemini, Claude, and Copilot are being incorporated rapidly into higher education and health professions training [1–3]. Unlike traditional rule-based systems, GenAI models can generate new text, images, and conversational responses and can support interactive learning, personalized explanations, and formative feedback. These capabilities are especially relevant to health science education, where learners must develop clinical reasoning, reflective judgment, and informed decision-making [4,5].

The educational value of GenAI depends on how it is integrated into teaching and learning. Health science education has traditionally emphasized experiential learning, reflection, teamwork, and independent problem-solving as foundations of professional competence [7–9]. Critical thinking, diagnostic reasoning, and clinical judgment are central to safe practice. GenAI may support these processes when used as a scaffold, but it may also reduce authentic cognitive engagement when it replaces rather than supports learner reasoning [6,10,11].

Evidence suggests that GenAI can support metacognitive development through iterative feedback, simulated clinical scenarios, and interactive learning environments [12,13]. AI-supported learning has also been associated with improved motivation, engagement, and self-efficacy when used to supplement human instruction [14]. In medical and nursing education, AI-supported simulations may improve reasoning accuracy, reflection, and confidence [15,16]. Personalized AI tutoring may also increase learning efficiency by adapting task difficulty and feedback to learner needs [17].

At the same time, inappropriate or excessive AI use may weaken core cognitive skills [18,19]. Greater reliance on AI-generated answers can encourage cognitive offloading and reduce productive struggle, independent problem-solving, and careful verification [19–21]. Concerns about hallucinations, bias, and variable accuracy are also important in health science education because incorrect or oversimplified outputs may influence learners' reasoning and judgment [22].

As GenAI becomes more common in teaching, assessment, and simulation, evidence is needed to clarify its effects on cognitive outcomes. Existing studies report inconsistent findings: some indicate improvements in critical thinking and decision-making [10,30–32], whereas others raise concerns about over-reliance, reduced independent reasoning, and weaker problem-solving [33–35]. These mixed findings make it difficult to determine whether GenAI consistently improves higher-order cognition or whether effects depend on the outcome being assessed and the way the technology is used.

Health science learners must develop both cognitive accuracy and ethical, clinically relevant judgment [36,37]. Educators therefore need approaches that help students use GenAI critically while preserving independent reasoning [36,38]. Curriculum design should support transparent, responsible use of AI while ensuring that students continue to practice core cognitive processes without unnecessary dependence on automated assistance [39,40].

Interventional findings remain inconsistent, and many individual studies are too small to estimate cognitive effects precisely. A quantitative synthesis is therefore needed to estimate pooled effects and determine whether effects differ across critical thinking, problem-solving, clinical reasoning, and decision-making. This meta-analysis focuses on interventional studies that provide sufficient quantitative information for standardized effect-size calculation.

Objective: To quantify the pooled effects of GenAI-supported educational interventions on critical thinking, problem-solving, clinical reasoning, and decision-making among health science learners.

METHODS

Evidence identification and quantitative eligibility

This meta-analysis was conducted within the evidence base identified under a prospectively registered protocol (PROSPERO CRD420251128495). The literature search was conducted by an information specialist in PubMed, Scopus, the Cochrane Library, and gray-literature sources, including Google Scholar, between September and November 2025. Reference lists of potentially eligible articles were also screened.

Medical Subject Headings (MeSH), keywords, and free-text terms were combined using Boolean operators. The search included terms for GenAI and large language models (for example, "generative AI," "GenAI," "large language model," "LLM," "ChatGPT," "GPT," "Bard," "Claude," and "AI assistant"), higher-order cognitive outcomes (including critical thinking, problem-solving, decision-making, and clinical reasoning), and health science learners or educators. The full search strategy is provided in Supplementary Table 1.

Eligible studies were peer-reviewed interventional investigations involving health science students or educators, evaluating a GenAI-supported educational intervention, and reporting critical thinking, problem-solving, clinical reasoning, or decision-making outcomes. Randomized, non-randomized, and quasi-experimental designs were considered. For this revised analysis, GenAI was operationally defined as an AI system capable of generating novel text, images, or conversational content in response to user input; conventional non-generative diagnostic or tutoring systems were not considered GenAI. Studies also had to provide sufficient quantitative information to calculate an effect size, including group sample sizes and continuous outcome statistics or equivalent extractable data. Wu et al. [63], which used the CC-Cruiser conventional AI diagnostic platform, was therefore removed from the quantitative GenAI synthesis.

Data extraction

Data extraction was independently conducted by two authors (AAA and NKA), with disagreements resolved by discussion and consensus. A standardized Microsoft Excel form captured author, year, country, study design, discipline, learner level, sample size, setting, AI tool, intervention, comparator, outcome domain, assessment time point, group means or mean changes, standard deviations, sample sizes, calculated effect size and variance, and risk-of-bias judgments.

Risk-of-bias assessment

Risk of bias was assessed using design-appropriate established tools: Cochrane RoB 2 for randomized controlled trials, ROBINS-I for non-randomized interventions, and Joanna Briggs Institute criteria for quasi-experimental studies. Assessments were performed independently by two reviewers, with disagreements resolved through discussion.

Statistical analysis

Extracted quantitative data were entered into Microsoft Excel and analyzed in Python 3.11.5. Standardized mean differences were calculated as Hedges' g with 95% confidence intervals. For pre/post designs, change scores were used when available; otherwise, post-intervention means were used. To reproduce the original analytical approach after exclusion of Wu et al., DerSimonian-Laird random-effects models were used. Statistical heterogeneity was evaluated using Cochran's Q, I^2, and τ^2. Outcome-domain subgroup analyses were performed for critical thinking, problem-solving, clinical reasoning, and decision-making. Fifteen unique studies contributed 21 outcome-domain effect estimates because some studies reported more than one eligible cognitive domain. These effect estimates were analyzed separately in the reconstructed overall model; consequently, analyses that use all 21 estimates should be interpreted cautiously because estimates originating from the same study are not fully independent. Funnel plots and Begg's and Egger's tests were used to explore small-study effects, and leave-one-effect-estimate-out sensitivity analyses assessed the influence of individual effect estimates. Given the substantial heterogeneity, within-study dependence, and small numbers in some subgroups, publication-bias and subgroup findings were interpreted cautiously.

Ethics

Ethical approval was not required because the analysis used data from publicly available published studies.

Natural Resources for Human Health
eISSN: 2583-1194

RESULTS

Quantitative evidence base

Fifteen unique interventional studies provided sufficiently complete and comparable quantitative data for effect-size calculation after exclusion of Wu et al. [63], whose CC-Cruiser intervention did not meet the operational definition of GenAI. The 15 studies contributed 21 outcome-domain effect estimates because several studies reported more than one eligible cognitive outcome. Studies that lacked means, standard deviations, sample sizes, or otherwise extractable continuous outcome data were not included in the quantitative synthesis.

Characteristics of studies contributing to the meta-analysis

Author	Year	Country	Design	Discipline	N	AI tool	Comparator	Outcome domain
Ahn et al. [68]	2025	Korea	Randomized Controlled Trial	Medicine	118	LLM-based diagnostic CDSS	No AI support	Decision-making
Akutay et al. [64]	2024	Turkey	Randomized controlled trial	Nursing	188	Gen-AI	Non-AI created case	Problem solving
Arkan et al.[26]	2025	Turkey	Randomized Controlled Trial	Nursing	96	ChatGPT (GPT-3.5)	Traditional nursing process education	Problem-Solving
Brügge et al.[47]	2024	Germany	Randomized Controlled Trial	Medicine	21	ChatGPT-3.5	AI-simulated patient conversation only	Clinical reasoning / Decision-making
Çiçek et al.[65]	2025	Turkiye	Randomized controlled trial	Medicine	115	ChatGPT-3.5	Expert-written feedback	Clinical reasoning
Haupt et al.[48]	2025	Germany	Randomized Controlled Trial	Dentistry	59	ChatGPT-4o	No support tool (control)	Cognitive decision-making
Huang et al. [73]	2025	China	Randomized controlled trial	Medicine	56	Custom GPT-based simulated patient	Traditional instructor-led role-playing	Clinical reasoning / problem-solving / decision-making
Jiang et al.[57]	2024	China	Randomized Controlled Trial	Medicine	61	Chatbot-based Standardized Patient (CSP)	Traditional SP-CBL	Clinical reasoning
Kejingyun et al.[58]	2025	China	Randomized controlled trial	Nursing	100	ChatGPT (LLM-integrated PBL)	Traditional Problem-Based Learning	Critical thinking

Sahin et al.[75]	2025	Turkiye	Randomized controlled trial	physiotherapy	94	ChatGPT	traditional instructor-led PBL approach	Problem solving, Critical thinking and decision making
Tabuchi et al.[49]	2024	Japan / UK	Randomized controlled trial	Optometry / Vision Science	55	Stable Diffusion XL (LoRA-optimized GenAI image generator)	Traditional 10-minute educational video	Decision-making
Tyrrell et al.[50]	2025	United Kingdom	Randomized crossover trial	Medicine	378	Sim Converse AI virtual patient simulator	Actor-based consultation skills training	Decision making
Wang et al.[62]	2025	China	Randomized Controlled Trial	Medical Education	56	Custom GPT model based on ChatGPT GPTs platform	Traditional instructor-led role-playing	Composite clinical competence (reasoning, communication, history-taking)
Saatci et al. [66]	2024	Turkey	Randomized controlled trial	Nursing	180	ChatGPT; Gemini; Copilot; Bing; Canva	Traditional resources only	Problem solving
Nissen et al. [51]	2025	Germany	Randomized cross-over trial	Medicine	154	ChatGPT-4 (GPT-4-turbo-2024-04-09)	Generic expert feedback	Clinical reasoning

Risk of bias

The risk-of-bias assessments indicated that most randomized trials had low risk or some concerns, commonly related to allocation concealment, blinding limitations inherent to educational interventions, and incomplete reporting of deviations from intended interventions. Non-randomized and quasi-experimental studies were more susceptible to confounding and selection bias. These limitations were considered when interpreting pooled effect magnitudes.

Overall pooled effect

The revised random-effects meta-analysis showed a statistically significant overall effect favoring GenAI-supported educational interventions (Hedges' $g = 0.58$, 95% CI 0.22–0.94, $p = 0.002$). Heterogeneity remained considerable ($Q = 475.07$, $p < 0.001$; $I^2 = 95.79\%$; $\tau^2 = 0.651$), indicating substantial variability across effect estimates and cognitive domains. The overall estimate is based on 21 outcome-domain effect estimates from 15 unique studies and should therefore be interpreted with caution because some studies contributed more than one estimate.

Natural Resources for Human Health
eISSN: 2583-1194

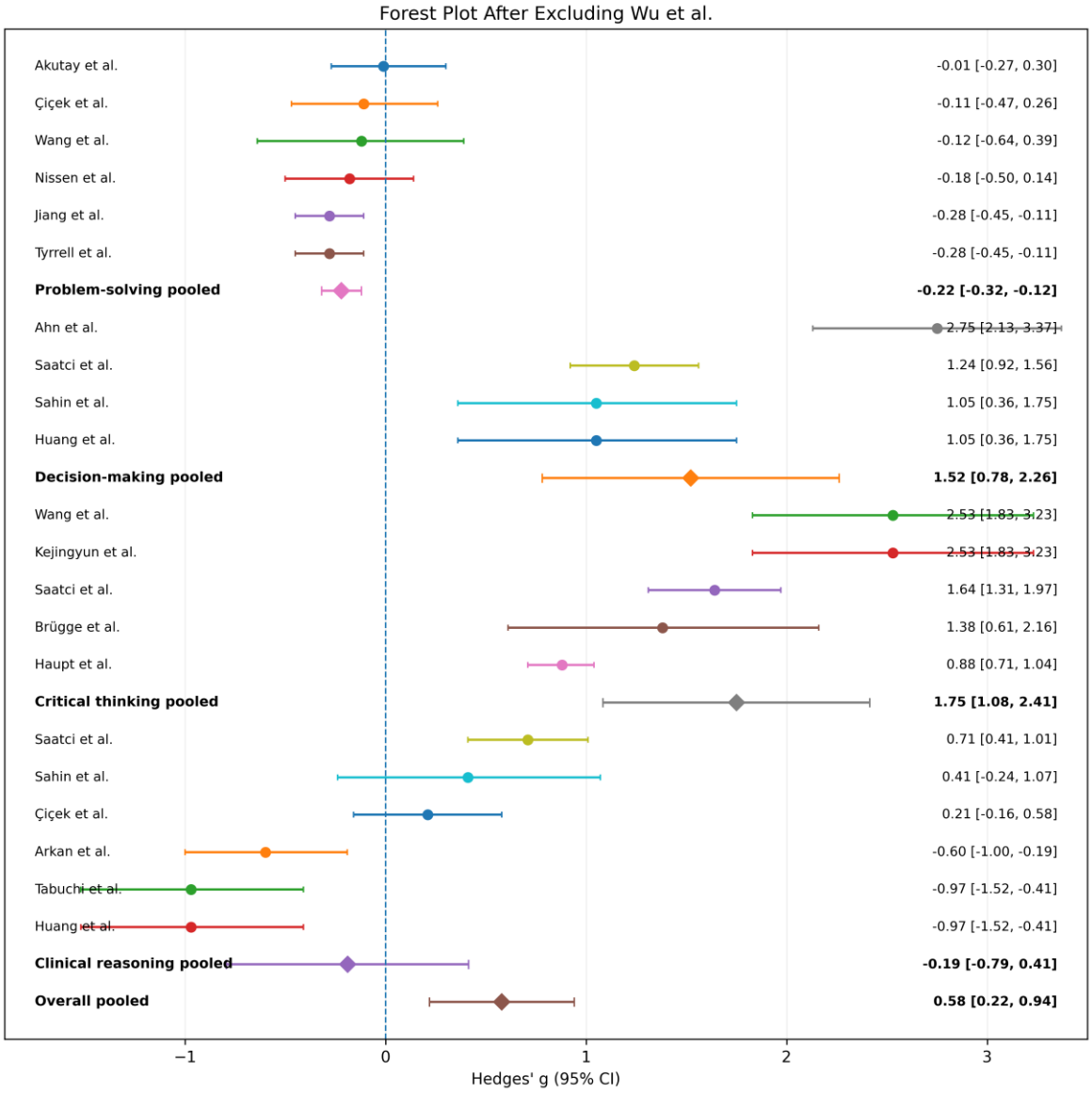

Figure 1. Forest plot of revised overall and outcome-domain effects after exclusion of Wu et al. Fifteen unique studies contributed 21 outcome-domain effect estimates.

Outcome-domain analyses

Effects varied markedly by cognitive domain. After removal of Wu et al., decision-making demonstrated a large positive random-effects estimate (g = 1.52, 95% CI 0.78–2.26; I^2 = 85.76%). Critical thinking also showed a large positive pooled effect (g = 1.75, 95% CI 1.08–2.41; I^2 = 91.82%). In contrast, problem-solving showed a small negative pooled effect (g = −0.22, 95% CI −0.32 to −0.12; I^2 = 0%), while clinical reasoning was not statistically significant (g = −0.19, 95% CI −0.79 to 0.41; I^2 = 91.09%). These differences indicate that the overall pooled effect should not be interpreted as uniform across cognitive outcomes.

Publication-bias assessment

Visual inspection of the revised funnel plot suggested some asymmetry. Begg's rank-correlation test was not statistically significant (Kendall's τ = 0.277, approximate Z = 1.73, p = 0.084), and Egger's regression test likewise did not provide statistically significant evidence of small-study effects (intercept = 3.07, t = 1.41, p = 0.176). These results should not be interpreted as evidence that publication bias is absent because heterogeneity was very high, some studies contributed multiple effect estimates, and the tests have limited reliability under these conditions.

Natural Resources for Human Health
eISSN: 2583-1194

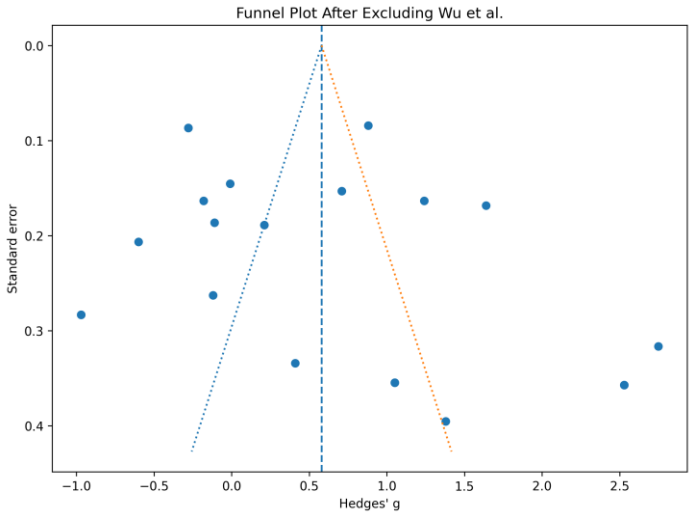

Figure 2. Funnel plot for the revised overall meta-analysis after exclusion of Wu et al.

Sensitivity analysis

Leave-one-effect-estimate-out sensitivity analysis showed that the pooled result remained positive after sequential omission of each of the 21 effect estimates. Random-effects estimates ranged from g = 0.47 to 0.65. The smallest pooled estimate occurred after removal of the Ahn et al. decision-making effect (g = 0.47, 95% CI 0.12–0.82), whereas the largest occurred after removal of the Tabuchi et al. or Huang et al. clinical-reasoning effect (g = 0.65, 95% CI 0.29–1.02). Heterogeneity remained very high across all leave-one-out models (I^2 approximately 95.35%–96.00%), indicating that no single effect estimate accounted for the observed variability.

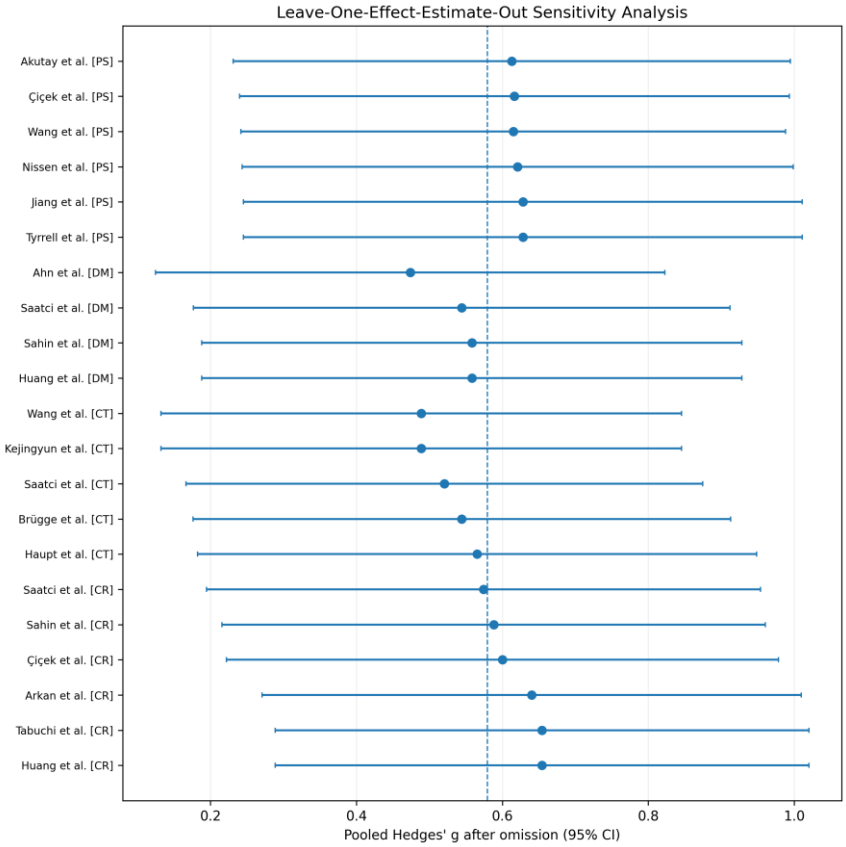

Figure 3. Leave-one-effect-estimate-out sensitivity analysis after exclusion of Wu et al.

DISCUSSION

This revised meta-analysis quantified the effects of GenAI-supported educational interventions on higher-order cognitive outcomes in health science education after removing a conventional non-generative AI study from the quantitative synthesis. The overall random-effects estimate remained positive and statistically significant, but the magnitude and direction of effects differed substantially across domains. Decision-making and critical thinking showed the strongest positive pooled effects, clinical reasoning was not statistically significant, and problem-solving showed a small negative pooled effect.

Overall cognitive effects of GenAI in health science education

The moderate overall pooled effect is consistent with evidence that GenAI can support learning through adaptive feedback, interactive assistance, simulation, and personalized educational support [2–4]. Broader higher education research also indicates that perceived usefulness, learner behavior, and institutional readiness influence the effectiveness of GenAI-supported learning [1,22,32]. Reviews in medical and health professions education describe potential benefits such as scalable tutoring and simulation, while also identifying risks related to inaccuracy, over-reliance, and ethical use [5,13,17,38].

The overall effect should nevertheless be interpreted cautiously. Heterogeneity remained extremely high after removal of Wu et al., showing that the variability cannot be explained by that study alone. Differences in AI tools, pedagogical designs, learner populations, comparators, outcome instruments, and assessment timing are likely contributors [3,4,6,11,12]. In addition, several studies contributed more than one cognitive-domain effect estimate to the reconstructed overall analysis, creating within-study dependence that may affect the precision of the pooled estimate. For these reasons, domain-specific results are more informative than the overall summary alone.

Effects on decision making

Decision-making showed the largest revised pooled effect after removal of Wu et al. This may reflect a close match between GenAI capabilities and decision-support learning tasks. GenAI systems can structure alternatives, provide immediate feedback, and support comparison of possible actions, all of which may help learners practice judgment in case-based and simulated settings [5,13,14,25,47,48,50].

Evidence from educational and clinical research suggests that structured decision-support interventions can improve the quality of judgment [27]. The present findings are consistent with the view that GenAI may support decision-making when it is used to prompt evaluation and reflection rather than replace learner judgment. The substantial heterogeneity within this subgroup, however, indicates that the size of benefit varies across interventions and contexts.

Effects on critical thinking

Critical thinking also showed a large positive pooled effect. Critical thinking develops through reflection, evaluation, comparison of evidence, and feedback rather than passive receipt of information [8,9]. GenAI can support these processes when learners are asked to compare perspectives, challenge assumptions, verify outputs, and justify conclusions [2–4,10]. Empirical work also suggests that GenAI may support knowledge construction and self-efficacy, which could contribute to improved critical-thinking performance [10,21,31].

The benefit of GenAI for critical thinking is therefore likely to depend on instructional design. When learners are required to critique, verify, or explain AI-generated outputs, GenAI can function as a scaffold for analytical engagement [20]. When the technology provides direct answers without structured reflection or verification, these benefits may be reduced [3,6].

Clinical reasoning outcomes

Clinical reasoning did not show a statistically significant pooled effect. Clinical reasoning requires pattern recognition, contextual interpretation, integration of evidence, and experience-based judgment [7,8]. GenAI can provide plausible explanations and diagnostic suggestions, but those outputs do not necessarily strengthen the learner's own reasoning process, particularly when automated responses replace active diagnostic synthesis.

Previous research has cautioned that GenAI can produce fluent but inaccurate or biased reasoning, which may mislead learners who are not prepared to evaluate AI outputs critically [2,5,13,17,40]. Variation in intervention design and outcome measurement may also explain the inconsistent clinical-reasoning findings. Structured reflection, explanation, and justification tasks may therefore be necessary when GenAI is used to support clinical reasoning [20].

Negative effect on problem solving

The small but statistically significant negative pooled effect on problem-solving warrants attention. Problem-solving requires learners to define problems, generate hypotheses, test strategies, and reflect on errors. These processes may be weakened when GenAI supplies direct solutions or step-by-step answers before learners have engaged independently [8,20]. Research on cognitive offloading similarly suggests that poorly structured GenAI use can reduce productive struggle and independent engagement [18–20,33–35].

Evidence from higher education raises similar concerns: GenAI may improve efficiency while impairing learning when it substitutes for active problem-solving rather than supporting it [33,35]. In health science education, where independent problem-solving is essential for safe practice, instructional designs should preserve learner responsibility, require justification of answers, and include opportunities to solve cases without AI assistance.

Educational, faculty, and policy implications

The findings support outcome-specific rather than uniform GenAI integration. GenAI appears most promising for decision-making and critical thinking, while clinical reasoning and problem-solving require more careful design, supervision, and evaluation. Faculty attitudes, institutional policies, and ethical frameworks also influence how GenAI is used in practice [6,15,36]. Clear expectations regarding transparency, verification, and appropriate reliance are needed to protect core cognitive competencies [37,40].

From a curriculum perspective, GenAI should be integrated with experiential and reflective learning cycles [8], explicit critical-thinking goals [9], and assessments that evaluate reasoning processes rather than final answers alone [3,4]. Educators should also consider staged use, in which students first attempt a task independently and then use GenAI for comparison, feedback, or reflection.

Strengths and limitations

A key strength of this meta-analysis is its focus on interventional evidence and its use of standardized effect sizes, random-effects synthesis, risk-of-bias assessment, small-study-effect exploration, and sensitivity analysis. Several limitations require emphasis. First, heterogeneity was very high, reflecting differences in AI tools, pedagogical designs, learner populations, comparators, outcome instruments, and assessment timing. Second, some cognitive-domain analyses included few effect estimates, reducing precision. Third, 15 unique studies contributed 21 outcome-domain effect estimates, and the reconstructed overall, funnel-plot, and leave-one-out analyses treated these estimates separately; because multiple estimates from the same study are correlated, standard errors may be underestimated and statistical tests should be interpreted cautiously. Fourth, combining different educational designs and outcome instruments through standardized mean differences improves comparability but does not remove conceptual differences among outcomes. Fifth, most outcomes were measured over relatively short periods, limiting conclusions about durable cognitive development. Finally, the search was completed in November 2025, so later studies were not captured. These limitations support placing greater interpretive weight on the domain-specific findings than on the single overall pooled estimate.

CONCLUSION

GenAI-supported educational interventions were associated with a moderate overall improvement in higher-order cognitive outcomes in health science education after exclusion of a conventional non-generative AI study. Effects were strongly domain dependent: decision-making and critical thinking showed large positive pooled effects, clinical reasoning showed no statistically significant pooled benefit, and problem-solving showed a small negative pooled effect. Very high heterogeneity and within-study dependence among multiple outcome estimates limit the certainty of the overall summary. Future trials should use clearly defined GenAI interventions, standardized cognitive outcome measures, transparent protocols, appropriate comparators, and longer follow-up, and future meta-analyses should use methods that explicitly account for correlated outcomes contributed by the same study.

DECLARATIONS

Ethics approval and consent to participate

This study used data from published articles; ethical approval and participant consent were therefore not required.

Acknowledgments

We thank the authors and participants of the original studies contributing data to this meta-analysis.

Availability of data and materials

This meta-analysis used publicly available data from published articles. Relevant data used in the analyses are reported in the manuscript.

Funding

The authors received no specific funding for this work.

Consent for publication

Not applicable.

Competing interests

The authors declare that no competing interests exist.

Author contributions

AAA conceived of and designed the study, analyzed and interpreted the results, and wrote the paper. NKA made significant contributions in all these areas; participated in drafting, revising, or critically reviewing the article; and agreed to be accountable for all aspects of the work. All authors read and approved the final manuscript

Abbreviations

AI, Artificial Intelligence; GenAI, Generative Artificial Intelligence; JBI, Joanna Briggs Institute; LLM, Large Language Model; RCT, Randomized Controlled Trial; SMD, Standardized Mean Difference; CI, Confidence Interval.

REFERENCES

[1] Jain, K.K. and J. Raghuram, Gen-AI integration in higher education: Predicting intentions using SEM-ANN approach. Education and Information Technologies, 2024. 29(13): p. 17169-17209.
[2] Kasneci, E., et al., ChatGPT for good? On opportunities and challenges of large language models for education. Learning and individual differences, 2023. 103: p. 102274.
[3] Pratschke, B.M., Generative AI and education: Digital pedagogies, teaching innovation and learning design. 2024: Springer.
[4] Bozkurt, A., Unleashing the potential of generative AI, conversational agents and chatbots in educational praxis: A systematic review and bibliometric analysis of GenAI in education. 2023, International Council for Open and Distance Education Oslo, Norway. p. 261-270.
[5] Cheng, Y. and L. Zhu, A review of ChatGPT in medical education: exploring advantages and limitations. International Journal of Surgery, 2025: p. 10.1097.
[6] Cespedes, A.A., Pedagogical shifts in the age of GenAI: Faculty perspectives from a higher education context. Journal of Pedagogical Sociology and Psychology, 2025. 7(3): p. 35-48.
[7] Benner, P., et al., Educating nurses: A call for radical transformation. 2009: John Wiley & Sons.
[8] Kolb, D.A., Experiential learning: Experience as the source of learning and development. 2014: FT press.
[9] Facione, P.A., Critical thinking: What it is and why it counts. Insight assessment, 2011. 1(1): p. 1-23.
[10] He, S. and Y. Lu, The effectiveness of Gen AI in assisting students' knowledge construction in humanities and social sciences courses: learning behaviour analysis. Interactive Learning Environments, 2024. 32(10): p. 7041-7062.
[11] Mengesha, Z., et al., Integrating generative AI into public health education: A mixed methods evaluation of student experiences. 2025.

[12] Batista, J., A. Mesquita, and G. Carnaz, Generative AI and higher education: Trends, challenges, and future directions from a systematic literature review. Information, 2024. 15(11): p. 676.
[13] Preiksaitis, C. and C. Rose, Opportunities, challenges, and future directions of generative artificial intelligence in medical education: scoping review. JMIR medical education, 2023. 9: p. e48785.
[14] Fung, T.C.J., et al., Effects of generative artificial intelligence (GenAI) patient simulation on perceived clinical competency among global nursing undergraduates: a cross-over randomised controlled trial. BMC nursing, 2025. 24(1): p. 934.
[15] Ichikawa, T., et al., Generative Artificial Intelligence in Medical Education—Policies and Training at US Osteopathic Medical Schools: Descriptive Cross-Sectional Survey. JMIR Medical Education, 2025. 11(1): p. e58766.
[16] Fung, T.C.J., et al., Effects of generative artificial intelligence (GenAI) patient simulation on clinical competency among global nursing undergraduates: A cross-over randomised controlled trial. 2025.
[17] Garcia, M.B., et al., Teaching Medicine with Generative Artificial Intelligence (GenAI): A Review of Practices, Pitfalls, and Possibilities in Medical Education. Teaching in the Age of Medical Technology, 2025: p. 123-156.
[18] Daniel, K., M.M. Msambwa, and Z. Wen, Can Generative AI Revolutionise Academic Skills Development in Higher Education? A Systematic Literature Review. European Journal of Education, 2025. 60(1): p. e70036.
[19] Xia, Q., et al., The impact of generative AI on university students' learning outcomes via Bloom's taxonomy: a meta-analysis and pattern mining approach. Asia Pacific Journal of Education, 2025: p. 1-31.
[20] Yan, L., et al., Beyond efficiency: Empirical insights on generative AI's impact on cognition, metacognition and epistemic agency in learning. 2025, Wiley Online Library. p. 1675-1685.
[21] Kim, M., Impact of Information Quality Perception of Generative AI on Critical Thinking: The Mediating Roles of Recommendation Behavior and Self-Efficacy. Asia-pacific Journal of Convergent Research Interchange (APJCRI), 2025: p. 57-68.
[22] Hsiao, C.-H. and K.-Y. Tang, Beyond acceptance: an empirical investigation of technological, ethical, social, and individual determinants of GenAI-supported learning in higher education. Education and Information Technologies, 2025. 30(8): p. 10725-10750.
[23] Zheng, J., et al. Evaluating the Impact of Using GenAI in Higher Education for University Students. in 2025 5th International Conference on Artificial Intelligence and Education (ICAIE). 2025. IEEE.
[24] Malik, T., et al. Exploring the transformative impact of generative AI on higher education. in Conference on e-Business, e-Services and e-Society. 2023. Springer.
[25] Ahn, S., et al., Impact of an Evidence-Based Large Language Model (LLM) Diagnostic Decision Support System: a Randomised Controlled Trial. Studies in health technology and informatics, 2025. 329: p. 273-277.
[26] Arkan, B., Ö. Dalli, and B. Varol, The impact of ChatGPT training in the nursing process on nursing students' problem-solving skills, attitudes towards artificial intelligence, competency, and satisfaction levels: single-blind randomized controlled study. Nurse education today, 2025. 152: p. 106765.
[27] Branda, M.E., et al., Shared Decision-Making for Patients Hospitalized with Acute Myocardial Infarction: a Randomized Trial. Patient preference and adherence, 2022. 16: p. 1395-1404.
[28] Brown, C. and K. Muir, EVALUATION OF AN ARTIFICIAL INTELLIGENCE (AI) ENHANCED APPLICATION FOR HYPERACUTE STROKE USING CLINICAL SIMULATION. European stroke journal, 2023. 8(2): p. 133.
[29] Chung, S., et al., 407 IMPACT OF ARTIFICIAL INTELLIGENCE SYSTEMS FOR UPPER GASTROINTESTINAL BLEEDING ON CLINICIAN TRUST AND LEARNING USING LARGE LANGUAGE MODELS: a RANDOMIZED PILOT SIMULATION STUDY. Gastroenterology, 2024. 166(5): p. S-95-S-96.
[30] Fu, Q., C.-P. Low, and K. Loh. Exploring the Impact of Gen-AI-Enabled Gamification on Student Motivation, Engagement, and Learning Outcomes. in Conference Proceedings. The Future of Education 2025. 2025.
[31] Qi, J., J.a. Liu, and Y. Xu, The role of individual capabilities in maximizing the benefits for students using GenAI tools in higher education. Behavioral Sciences, 2025. 15(3): p. 328.
[32] Alotaibi, S.M.F., Determinants of Generative Artificial Intelligence (GenAI) adoption among university students and its impact on academic performance: the mediating role of trust in technology. Interactive Learning Environments, 2025: p. 1-30.

[33] Gökoğlu, S. and F. Erdoğdu, The effects of GenAI on learning performance: A meta-analysis study. Educational Technology & Society, 2025. 28(3).
[34] Ng, J., et al., Exploring Students' Perceptions and Satisfaction of Using GenAI-ChatGPT Tools for Learning in Higher Education: A Mixed Methods Study. SN Computer Science, 2025. 6(5): p. 476.
[35] Macko, V., et al. Augmenting Active Learning with GenAI: Enhancement or Impairment? Evidence from a Data Visualization Course. in European Conference on Technology Enhanced Learning. 2025. Springer.
[36] Dong, C., et al., Generative AI in Healthcare: Insights from Health Professions Educators and Students. International Medical Education, 2025. 4(2): p. 11.
[37] Sallam, M. and M. Sallam, Ethical aspects of implementing generative artificial intelligence in medical education: a narrative review. History and Philosophy of Medicine, 2025. 7: p. 18-25.
[38] Pham, T.D., et al., The impact of generative AI on health professional education: A systematic review in the context of student learning. Medical Education, 2025.
[39] Jongkind, R., et al. Studying with GenAI: Cross-sectional study on usage patterns, needs, competencies and ethical perspectives of medical informatics students. in Frontiers in Education. 2025. Frontiers.
[40] Ning, Y., et al., Generative artificial intelligence and ethical considerations in health care: a scoping review and ethics checklist. The Lancet Digital Health, 2024. 6(11): p. e848-e856.
[41] Moher, D., et al., Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Systematic reviews, 2015. 4: p. 1-9.
[42] Kloda, L.A., J.T. Boruff, and A.S. Cavalcante, A comparison of patient, intervention, comparison, outcome (PICO) to a new, alternative clinical question framework for search skills, search results, and self-efficacy: a randomized controlled trial. Journal of the Medical Library Association: JMLA, 2020. 108(2): p. 185.
[43] Eldridge, S., et al., Revised Cochrane risk of bias tool for randomized trials (RoB 2.0): additional considerations for cluster-randomized trials. Cochrane Methods Cochrane Database Syst Rev, 2016. 10(suppl 1).
[44] Jüni, P., et al., Risk of bias in non-randomized studies of interventions (ROBINS-I): detailed guidance. Br Med J, 2016. 355: p. i4919.
[45] Maheshwarappa, H.M. and S. Majumder, Critical Appraisal of Randomized Controlled Trials: An Overview. Indian Journal of Respiratory Care, 2023. 12(2): p. 164.
[46] Institute, J.B. CRITICAL APPRAISAL TOOLS. [cited 2021 December 15]; Available from: https://jbi.global/critical-appraisal-tools.
[47] Brügge, E., et al., Large language models improve clinical decision making of medical students through patient simulation and structured feedback: a randomized controlled trial. BMC medical education, 2024. 24(1): p. 1391.
[48] Haupt, F., T. Rödig, and P. Liersch, Evaluating ChatGPT-4o as an Educational Support Tool for the Emergency Management of Dental Trauma: Randomized Controlled Study Among Students. JMIR Med Educ, 2025. 11: p. e80576.
[49] Tabuchi, H., et al., Comparative educational effectiveness of AI generated images and traditional lectures for diagnosing chalazion and sebaceous carcinoma. Scientific reports, 2024. 14(1): p. 29200.
[50] Tyrrell, E.G., et al., Web-Based AI-Driven Virtual Patient Simulator Versus Actor-Based Simulation for Teaching Consultation Skills: Multicenter Randomized Crossover Study. JMIR Formative Research, 2025. 9: p. e71667.
[51] Nissen, L., et al., A randomised cross-over trial assessing the impact of AI-generated individual feedback on written online assignments for medical students. Medical Teacher, 2025: p. 1-7.
[52] Chen, Y., Evaluation of the impact of AI-driven personalized learning platform on medical students' learning performance. Front Med (Lausanne), 2025. 12: p. 1610012.
[53] Cheng, C.T., et al., Artificial intelligence-based education assists medical students' interpretation of hip fracture. Insights Imaging, 2020. 11(1): p. 119.
[54] Fung, T.C.J., et al., Effects of generative artificial intelligence (GenAI) patient simulation on perceived clinical competency among global nursing undergraduates: a cross-over randomised controlled trial. BMC Nurs, 2025. 24(1): p. 934.
[55] Hou, J., et al., Application of DeepSeek-assisted problem-based learning in hematology residency training. BMC Medical Education, 2025. 25(1).
[56] Huang, S., et al., Exploring the Application Capability of ChatGPT as an Instructor in Skills Education for Dental Medical Students: Randomized Controlled Trial. J Med Internet Res, 2025. 27: p. e68538.

[57] Jiang, Y., et al., Enhancing medical education with chatbots: a randomized controlled trial on standardized patients for colorectal cancer. BMC Medical Education, 2024. 24(1): p. 1511.
[58] Kejingyun, S. and R. Mingjun, Randomized Controlled Study on the Impact of Problem-Based Learning Combined With Large Language Models on Critical Thinking Skills in Nursing Students. Nurse educator, 2025. 50(4): p. 216-220.
[59] Li, L., et al., The role of generative AI tools in case-based learning and teaching evaluation of medical biochemistry. BMC Medical Education, 2025. 25(1).
[60] Shin, H., et al., The Impact of Artificial Intelligence-Assisted Learning on Nursing Students' Ethical Decision-making and Clinical Reasoning in Pediatric Care: A Quasi-Experimental Study. Comput Inform Nurs, 2024. 42(10): p. 704-711.
[61] Song, D., et al., 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, 2025: p. 104549.
[62] Wang, Z., et al., Feasibility study of using GPT for history-taking training in medical education: a randomized clinical trial. BMC medical education, 2025. 25(1): p. 1030.
[63] Wu, D., et al., Artificial intelligence-tutoring problem-based learning in ophthalmology clerkship. Annals of translational medicine, 2020. 8(11).
[64] Akutay, S., H. Yuceler Kacmaz, and H. Kahraman, The effect of artificial intelligence supported case analysis on nursing students' case management performance and satisfaction: A randomized controlled trial. Nurse Educ Pract, 2024. 80: p. 104142.
[65] Çiçek, F.E., et al., ChatGPT versus expert feedback on clinical reasoning questions and their effect on learning: a randomized controlled trial. Postgraduate medical journal, 2025. 101(1195): p. 458-463.
[66] Saatçi, G., S. Korkut, and A. Ünsal, The effect of the use of artificial intelligence in the preparation of patient education materials by nursing students on the understandability, actionability and quality of the material: A randomized controlled trial. Nurse Education in Practice, 2024. 81: p. 104186.
[67] Ali, M., S. Rehman, and E. Cheema, Impact of artificial intelligence on the academic performance and test anxiety of pharmacy students in objective structured clinical examination: a randomized controlled trial. International Journal of Clinical Pharmacy, 2025. 47(4): p. 1034-1041.
[68] Ahn, S., et al., Impact of an Evidence-Based Large Language Model (LLM) Diagnostic Decision Support System: A Randomised Controlled Trial. Stud Health Technol Inform, 2025. 329: p. 273-277.
[69] Fazlollahi, A.M., et al., Effect of Artificial Intelligence Tutoring vs Expert Instruction on Learning Simulated Surgical Skills Among Medical Students: A Randomized Clinical Trial. JAMA Netw Open, 2022. 5(2): p. e2149008.
[70] Giglio, B., et al., Artificial Intelligence-Augmented Human Instruction and Surgical Simulation Performance: A Randomized Clinical Trial. JAMA Surg, 2025. 160(9): p. 993-1003.
[71] Goh, E., et al., Large Language Model Influence on Diagnostic Reasoning: A Randomized Clinical Trial. JAMA Netw Open, 2024. 7(10): p. e2440969.
[72] Han, J.-W., J. Park, and H. Lee, Analysis of the effect of an artificial intelligence chatbot educational program on non-face-to-face classes: a quasi-experimental study. BMC Medical Education, 2022. 22(1): p. 830.
[73] Huang, Y., et al., Implementation and evaluation of an optimized surgical clerkship teaching model utilizing ChatGPT. BMC Med Educ, 2024. 24(1): p. 1540.
[74] Park, S.-A. and H.Y. Kim, 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, 2025. 31(2): p. 143-154.
[75] Ergezen Sahin, G., et al., Effects of artificial intelligence based physiotherapy educational approach in developing clinical reasoning skills: a randomized controlled trial. BMC Med Educ, 2025. 25(1): p. 1378.