Check for updates

Review Article

Effects of GenAI Interventions on Student Academic Performance: A Meta-Analysis

Journal of Educational Computing Research
2025, Vol. 0(0) 1–33
© The Author(s) 2025
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/07356331251349620
journals.sagepub.com/home/jec
Sage

Jiahe Gu and Zi Yan

Abstract

Generative artificial intelligence (GenAI) has the potential to change student learning. Despite the popularity of integrating this novel technology into teaching and learning practices, few meta-analyses have synthesised its effect in the education context with K-12 and college students. This review examined the effects of GenAI interventions on student academic performance. A total of 19 studies with 24 effect sizes were included. These studies either compared the GenAI group with control groups (n = 17, k = 22) or applied a repeated-measure design (n = 2, k = 2). The results revealed an overall large effect size (g = 0.683), supporting the arguments that GenAI can positively affect student academic achievement. Students with teacher support in the student-GenAI interaction have significantly larger gains (g = 1.426) than those without teacher support (g = 0.077). No other significant moderators were identified. We concluded by discussing the implications for policy and practice and provided suggestions for future research.

Keywords

GenAI, GenAI in education, meta-analysis, academic performance, teacher scaffolding, educational technology

The development of artificial intelligence (AI) and its integration into teaching and learning practices are transforming education. For instance, AI can facilitate student

Department of Curriculum and Instruction, The Education University of Hong Kong, Hong Kong, China

Corresponding Author:
Zi Yan, Department of Curriculum and Instruction, The Education University of Hong Kong, 10 Lo Ping Road, Tai Po, N.T., Hong Kong, China.
Email: zyan@eduhk.hk

learning by providing adaptive feedback (Liang et al., 2024) and visual performance reports (Liao et al., 2024). The launch of OpenAI's ChatGPT in November 2022 attracted public attention to one subset of AI, i.e., generative artificial intelligence (GenAI). GenAI can create new content based on generative models, making it distinctive from traditional non-generative models that focus on prediction, classification, or optimisation (Rashidi et al., 2024; Saish et al., 2025). Such a generative ability of GenAI tools resulted in a surge of global interest, with ChatGPT reaching more than one million users in just five days (Yu, 2023). Now, it has over 400 million weekly active users, according to OpenAI's spokesperson in February 2025.

In educational contexts, GenAI tools hold the potential to facilitate student learning. By engaging in realistic interactions with learners, GenAI tools not only provide answers from various disciplines, but also generate examples, recognise errors, and remember the context of the dialogue (Imran & Almusharraf, 2023; OpenAI, 2022). However, concerns have been raised about the over-reliance on GenAI, as it may lead to plagiarism and a decline in critical thinking ability (e.g., Lo et al., 2024). Extant empirical studies have also revealed conflicting results regarding the influence of GenAI on student academic achievement, with some studies revealing positive effects (e.g., Wu et al., 2024) and others reporting no effects (e.g., Escalante et al., 2023) or even negative effects (e.g., Niloy et al., 2024).

Given the rapid development of GenAI and mixed findings in existing research, there is a need to systematically synthesise the effects of GenAI interventions on student academic performance and explore potential moderating factors. As one of the first few meta-analytical reviews that narrowed the research scope from AI to GenAI, the current study aimed to (a) figure out the overall effects of GenAI interventions on student academic performance and (b) investigate the potential moderators that influence their effectiveness in educational settings. Following PRISMA guidelines, we examined four databases to determine the overall effect size. We also conducted sensitivity and moderator analyses to assess the robustness of the results and identify influential factors. The results can deepen the understanding of the effects of GenAI interventions on student learning outcomes and guide the effective use of GenAI tools in future educational practices.

What is GenAI?

GenAI is a type of AI technology that uses machine learning models "to learn the patterns and relationships in a dataset of human-created content" and "use the learned patterns to generate new content" (Google, 2023, How does generative AI work? Section, para. 1). More specifically, GenAI mainly uses a subset of machine learning, that is deep learning approach (Kalota, 2024; Strobel et al., 2024), to produce "previously unseen synthetic content, in any form and to support any task, through generative modelling" (Peñalvo & Ingelmo, 2023, p. 14).

Based on different generative models and techniques, GenAI can create new content in different forms (e.g., texts, images, and audio files) (Bengesi et al., 2024). For

instance, large language models (LLMs), which emerged in 2017, are primarily designed to process and generate texts (e.g., OpenAI's ChatGPT and Google's Bard) (Law, 2024). Unlike previous AI chatbots (e.g., Apple's Siri) (Kietzmann & Park, 2024), LLMs can analyse and summarise online content, generating new responses in a conversational format across various fields. This process resembles how humans produce novel texts from learned knowledge (Barrett & Pack, 2023). Beyond text generation, other GenAI tools like Midjourney for image creation and Sora for video production can also significantly impact education (Chiu, 2024; Liu et al., 2024).

Effects of GenAI Interventions on Student Academic Performance. GenAI interventions in this review refer to the integration of GenAI tools into teaching or learning practice with an aim to influence students' learning outcomes. The extant literature has summarised various tasks that GenAI could help with in educational settings. GenAI could serve multiple roles in enhancing teaching effectiveness, working as a "guide on the side" for content generation, a "co-designer" for curriculum development, and an "exploratorium" for assessment analysis and learning recommendations (Sabzalieva & Valentini, 2023). Beyond these functions, it could assist in lesson planning, student record tracking, course material translation, and student engagement (Alshraah et al., 2024). The integration of GenAI in teaching can increase teachers' working efficiency (Law, 2024), build teaching confidence (Cheah & Kim, 2025), enhance pedagogical competency (Alshraah et al., 2024), and motivate teachers to adopt innovative teaching and assessment methods (Bower et al., 2024). Such improvements in teaching practice could positively impact students' academic performance.

While the incorporation of GenAI into teaching practices plays an important role in students' learning, students' direct use of GenAI tools could exert a more straightforward influence on their academic performance. Grounded in constructivist theory and Vygotsky's concept of the Zone of Proximal Development (ZPD), GenAI tools can facilitate student learning by providing personalised feedback that adapts to individual needs, guiding learners progressively through their development (Coenen & Pfenninger, 2024; Zhang et al., 2024). GenAI-generated immediate and diverse feedback helps students understand their current performance, identify learning gaps, and formulate future goals (Xia et al., 2024; Yan, 2022). GenAI tools can work as a "study buddy" to facilitate student self-reflection, a "Socratic opponent" to develop argumentation skills, or a "collaboration coach" to facilitate group work (Sabzalieva & Valentini, 2023). Studies also revealed that GenAI interventions can increase students' motivation (e.g., Li, 2023; Song & Song, 2023) and self-efficacy (Teng, 2024) and decrease students' anxiety and embarrassment (Hsu et al., 2023). However, there are concerns about potential risks for academic development, such as the negative impact on critical thinking skills and "metacognitive laziness" (Fan et al., 2025; Susnjak & McIntosh, 2024). Students' use of AI to complete assignments may also diminish their motivation to develop skills, resulting in an educational crisis beyond academic dishonesty (Adeshola & Adepoju, 2023). Considering these impacts, this meta-analysis

specifically examines students' direct use of GenAI tools in their learning process, as understanding these effects is essential for informing educational practices.

A substantial body of empirical research has witnessed varied GenAI interventions among students, for instance, students receiving GenAI-generated feedback (Escalante et al., 2023) and engaging in formative assessment with GenAI-generated questions (Bachiri et al., 2023). However, the effectiveness of GenAI interventions on student academic performance varied across studies. Some studies found significant improvement in student academic achievement in the GenAI group (e.g., English writing, Liu & Xiao, 2024; Song & Song, 2023); some studies showed no significant differences between the GenAI and control groups (e.g., language learning, Escalante et al., 2023; and theoretical medical knowledge, Ba et al., 2024); and some studies revealed that students who used ChatGPT alone underperformed in mathematics than those who received teacher instruction (e.g., mathematics, Dasari et al., 2024). Such an inconclusive result reveals the need for a meta-analysis to systematically synthesise the overall effectiveness of GenAI interventions on student academic performance and the influencing factors that moderate the effects.

Previous Reviews. The extant reviews about GenAI have mainly explored this new technology through systematic review (e.g., Lo et al., 2024), scoping review (e.g., Preiksaitis & Rose, 2023), and bibliometric analysis (e.g., Bahroun et al., 2023). As for meta-analyses, scholars investigated users' perceptions of GenAI (e.g., Leiter et al., 2024) and the accuracy of GenAI's outputs in exams (e.g., the diagnostic performance of GenAI compared with physicians, Takita et al., 2024; medical responses, Wei et al., 2024). Despite these reviews providing insights into the trend, benefits, challenges, and performance of GenAI in educational settings, the comprehensive synthesis of the impact of GenAI interventions on academic performance has been less touched.

Only Sun and Zhou's (2024) meta-analysis has specifically explored the effectiveness of GenAI interventions on student academic performance. They found that GenAI can enhance student academic performance with a medium effect size (g = 0.533). They also found that students significantly improved their academic performance when GenAI generated texts, used in a collaborative learning approach, and included a sample size of 21–40. However, they only focused on a particular group of students (i.e., college students). They did not rigorously assess the quality of the included studies and did not specifically focus on peer-reviewed journal articles, which is vital in this fast-moving field.

Since GenAI is a subset of AI technology, meta-analytical reviews on the effects of broader AI technology could provide some valuable insights. Previous meta-analyses about AI interventions on student learning generally show a positive influence. Zheng et al. (2023) selected articles from 2001 to 2020 and found that AI technologies (e.g., expert systems or agent systems, natural language processing, and mixed technology) had a high effect size on learning achievement (g = 0.812). The moderator analysis revealed a significantly large effect size with a substantial sample size (more than 300), junior and senior high school students, engineering and technological science students,

AI being utilised in group settings, serving as policy-making advisors, and incorporating mixed hardware.

Wu and Yu (2024) performed a meta-analysis of 24 randomised studies to determine the effects of various types of AI chatbots on student learning outcomes, including machine learning-based chatbots, natural language processing-based chatbots, and hybrid chatbots. They also revealed that AI chatbots could significantly improve students' learning performances (d = 1.028). Students who received shorter interventions (i.e., lasting less than ten weeks) at the higher education level experienced a greater effect.

Apart from the studies that comprehensively analysed the effects of AI in education, there are studies mainly focused on one learning domain. For instance, Wang et al. (2024a) found that AI chatbots produced an overall positive effect on language learning performance (g = 0.484) compared to students who did not use chatbots. Four significant moderators were identified: educational level, learners' language levels, interface design (i.e., mobile-based interface vs. web-based interface), and interaction capability (i.e., chatbot-driven capability vs. user-driven capability). Other meta-analyses also revealed a positive effect of AI technology on student subject-specific learning (e.g., g = 0.351 for elementary students' mathematical learning, Hwang, 2022; d = 1.18 based on within-group samples; d = 0.39 based on 35 between-group samples for language learning, Lee & Lee, 2024; g = 0.343 for K-12 students' mathematical learning, Yi et al., 2024).

While Sun and Zhou's (2024) meta-analysis and previous reviews about AI identified several factors that may moderate the effects of GenAI interventions, other influencing factors, specifically regarding GenAI and research methodology, may also affect the impact. Instructors can directly apply general GenAI tools, designed by big companies, such as OpenAI's ChatGPT, or use course-specific GenAI tools tailored for their students. More specifically, course-specific GenAI tools could be developed through two methods: 1) fine-tuned text-to-text models designed for a specific context, which is based on a selected knowledge base (e.g., Bachiri et al., 2023), and 2) models based on existing LLMs like ChatGPT but modified by research teams to include additional learning functions, for instance, constructing assignment databases and learning profile databases within the ChatGPT-based learning system (e.g., Li, 2023). The direct application of the publicly accessible tool is convenient; however, the generated feedback is not content- or course-specific (Xia et al., 2024). Feedback research shows that students may experience greater learning benefits when they receive feedback that is concrete and specific to the content or course they are studying (Olivera-Aguilar et al., 2022; Shute, 2008). Hence, we examined the different effects between general and course-specific GenAI tools.

Apart from the technical component, the human element, specifically teacher support in classrooms where GenAI tools are used, is less explored in the existing research (Kizilcec et al., 2024). According to Tardy's (1985) social support framework, teacher support includes giving students informational, instrumental, emotional, and

appraisal support (Malecki & Demaray, 2002). When students use GenAI, teachers can provide ongoing support by giving advice, resources, and feedback. For instance, teachers share learning resources through GenAI-based learning systems (e.g., Baba et al., 2024) and give feedback on GenAI-generated content (e.g., Wu et al., 2024). However, in some studies, teachers let students use GenAI without instruction or only offer initial training rather than continuous support throughout the learning process. Teacher support can help facilitate the incorporation of appropriate GenAI outputs into students' work (Su et al., 2023); otherwise, students may lack the capacity to critically assess the quality of GenAI outputs. Therefore, we considered teacher support a potential moderator that can enhance the positive effects of using GenAI tools in education.

Regarding research design, the control groups used in the study for comparison matter. In some studies, students in the control group received teacher instruction or feedback; however, in other studies, students in the control group used other resources, such as Google search, online databases, or textbooks. Feedback information from different agents (e.g., human teachers and technology) may affect student learning differently (Panadero & Lipnevich, 2022). Thus, it is important to test the effects of the comparison groups—those receiving no feedback, teacher feedback, or feedback from other resources—on the effects of GenAI interventions.

The research method can affect the influence of educational interventions (McMillan et al., 2013). Hence, the methodological characteristics of the selected studies were carefully assessed through the study quality (i.e., study design, sampling method, random assignment, confounder report and control, data source, instrument source, and withdrawals and attrition rate) and included in this study as a potential moderator.

In sum, while previous meta-reviews revealed an overall positive effect of AI on academic performance, a comprehensive synthesis exclusively on GenAI remains scarce. Considering variant effect sizes in empirical studies, some factors may influence the effects of GenAI on academic performance. Based on both the data-driven and theory-driven evidence, the potential moderators of GenAI interventions were grouped into three categories: 1) implementation of GenAI, including *GenAI tool* (general GenAI tools or course-specific GenAI tools), *teacher support* (with or without support) and *intervention duration* (equal to or less than one month, or longer than one month); 2) the context of the study, including *educational level* (K-12 or higher education) and *discipline* (natural and applied science, or humanities and social science); and 3) research design, including *control group* (no feedback, teacher feedback, or feedback from other resources), *sample size* (small or large), and *study quality* (strong, moderate or weak).

Contributions of the Present Study. The current review aimed to provide a comprehensive review of the effects of GenAI interventions on academic performance. First, rather than including broad AI technology, we specifically explored the effects of GenAI on student academic achievement. After the release of ChatGPT sparked a wave of interest in GenAI among educational researchers and practitioners, GenAI has now been widely

discussed and used in the educational field. Therefore, it is essential to scrutinise the effects of this latest generation of AI technology on student academic achievement.

Second, we attempted to provide a comprehensive synthesis by including studies from different educational levels (i.e., K-12 and higher education) with different research designs (i.e., experimental studies, quasi-experimental studies, and studies with a repeated measures design). In this way, we could compare the influence of GenAI on students from different backgrounds and consider the influence of research design on the GenAI effects.

Third, our meta-analytical review carefully considered the quality of the included studies by only selecting peer-reviewed journal articles and critically assessing the quality of each selected study. Empirical studies with rigorous methodology are more likely to present robust evidence for the effectiveness of GenAI in this rapidly evolving field. The research questions (RQ) are listed below:

RQ 1: What is the overall effect of GenAI interventions on student academic performance?
RQ 2: What factors moderate the effects of GenAI interventions on student academic performance?

Method

Search Strategies and Databases

The literature search was conducted across four databases: ERIC, PsycINFO, Web of Science, and Scopus. These four databases were selected because they were comprehensive to include journal articles in the field of AI in education. Three groups of keywords about GenAI, feedback, and education were combined to form the search string: ("GAI" OR "Generative AI" OR "Generative Artificial Intelligence" OR "ChatGPT" OR "GPT" OR "Large Language Models" OR "LLM" OR "AlphaCode" OR "GitHub Copilot" OR "Bard") AND ("feedback" OR "assessment" OR "instruction" OR "scaffolding" OR "training") AND ("school" OR "education" OR "colleague*" OR "Tertiary Education" OR "higher education" OR "teacher*" OR "student*"). The synonyms for GenAI are modified from Bahroun et al.'s (2023) study, a bibliometric and content analysis on GenAI in education that used the search string including both GenAI and specific GenAI tools, such as ChatGPT and Bard. The search included papers published after 2017, considering that the transformer (the 'T' in GPT, generative pre-trained transformer) was first announced that year (Law, 2024).

Selection of Studies

The literature search was performed on 20 May 2024 using Title or Abstract, revealing 1938 results from four databases. A total of 1310 studies were left for screening after removing duplicates. An eligible paper has to meet the following criteria: (a) it

measured the effects of students using GenAI on student academic performance, (b) it adopted an experimental/quasi-experimental design comparing the GenAI group with a comparison group (no GenAI interventions, or an experimental group of non-AI feedback types, such as teacher feedback and peer feedback), or adopted a pre-post comparison design without a control group, and (c) it provided effect size or sufficient information to calculate effect size (e.g., means, standard deviations, and sample sizes), (d) it was written in English and published in a peer-review journal. This review excluded other types of literature, such as book chapters and dissertations.

Studies were excluded if (a) it investigated the effects of GenAI on non-academic outcomes, such as student perception (Kelly et al., 2023), critical thinking skills (Guo & Lee, 2023), and creativity (Habib et al., 2024); (b) it only compared the quality of GenAI-generated feedback and other feedback type but not investigated the comparative effectiveness on student academic performance (Almasre, 2024; Banihashem et al., 2024); (c) it explored the effects of teachers using GenAI during the teaching practice (e.g., preparing course materials) on student academic achievement (Ghafouri et al., 2024); (d) it was a theoretical paper (Bearman et al., 2024), review paper (Baber et al., 2023; Zirar, 2023), editorial opinion (Crawford et al., 2023), or personal reflection (Keath et al., 2024).

The selection process was guided by Page et al.'s (2021) PRISMA 2020 flow diagram (See Figure 1). The identified articles were closely examined through title and abstract in the first round of screening and full texts in the second round of screening based on the selection criteria. Around 10% of the initially identified articles were examined by two coders to ensure the reliability of the screening process. Inter-rater reliability (kappa) between the two coders was 0.48, indicating a moderate agreement (Fleiss, 1971). The disagreements were solved by discussing and consulting with an expert before proceeding to the next step. Finally, 19 studies with 24 effect sizes were selected for the meta-analysis.

Data Extraction

To enhance the consistency and reliability of the data extraction process, we developed a data extraction form tailored to the research questions. The form comprised five sections: (1) basic information about the studies, (2) implantation of GenAI, (3) context of the study, (4) research design, and (5) outcomes. The item was coded as missing data if the information was not reported in the selected studies.

The first section contained the basic information about the studies, such as title, author, journal, and publication years. The second section was the implantation of GenAI tools, including *GenAI tool*, *teacher support,* and *intervention duration*. The whole research period was coded as intervention duration because the specific time of using GenAI was seldom reported in the study. The third section relates to the context of the study, including *educational level* and *discipline*. The fourth section was the characteristics of the research design, including *control group*, *sample size*, and *study quality*. More specifically, the quality of the included studies was assessed against the

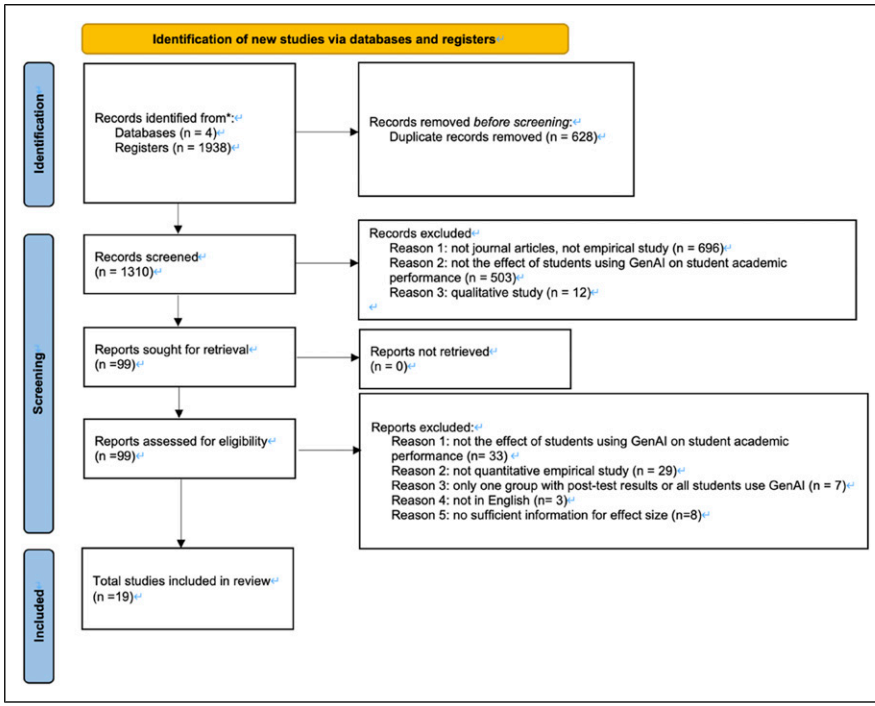

Figure 1. Prisma Flow Diagram for the Study Selection Process.

modified Effective Public Health Practice Project (EPHPP) instrument (Thomas et al., 2004), including five dimensions: study design, participant selection bias, confounder report and control, data collection methods, and withdrawals and dropouts. The "blinding" dimension in the original instrument was excluded in this meta-analysis because of the difficulty of blinding participants or researchers in educational research (Noetel et al., 2021). The detailed checklist for assigning a score (i.e., strong, moderate, or weak) to each dimension of the included studies was presented in the Supplemental Material. The final computing scores from the five dimensions were the overall quality of the included studies. The fifth section was outcome variables and estimates of effect size (i.e. sample size, mean, and standard deviation).

Statistical Analyses

Effect size calculation was conducted separately for studies with and without control groups. For the study with a control or comparison group, Cohen's *d* (Cohen, 1988) was used to calculate effect size. More specifically, for the studies with the control group and only post-test scores available, post-test scores were used in the formula; and for those with the control group and both pre- and post-test scores, the change scores between the

pre- and post-test results were used. As for those repeated-measure design studies without control groups, the effect sizes were calculated using Becker's (1988) formula. As few correlations have been reported in published articles, drawing from related studies (Borenstein et al., 2021), this study adopted a pre-post correlation of 0.5, a conventional practice commonly used in many meta-analyses (e.g., Yan et al., 2022; Zhan et al., 2023). All Cohen's *d* values were converted to Hedges' *g* (Hedges, 1981) to correct small sample bias. A positive effect size suggests better learning gains for the GenAI group compared with the control group or the positive effect after GenAI use in the single-group study.

Two- and three-level model comparison was conducted due to multiple effect sizes in 15.79% of the selected studies. Analyses of heterogeneity were performed by the I^2 test to figure out the degree of variance in effect sizes: 0%–40%, not important; 30%–60%, moderate; 50%–90%, substantial; and 75%–100% considerable (Shamseer et al., 2015). If heterogeneity were high, moderator analyses with a mixed-effects model would be conducted through meta-regression to identify the sources of variance. Three categories of potential moderators identified in the literature were tested, including the implementation of GenAI tools, the context of the study, and the research design.

To robust the findings of this review, the outliers were detected by checking whether effect sizes fell outside the range ($\overline{x}$– 3sd, $\overline{x}$+ 3sd) (Acuna & Rodriguez, 2004). The sensitivity analyses were also performed to test the influence on effect size when leaving out each study and excluding studies with weak quality. The existence of publication bias was checked through funnel plots (Light & Pillemer, 1984) and the three-level Egger regression test (Egger et al., 1997). If public bias existed, Vevea and Woods' (2005) selection model would be used to adjust the results. All data analyses were performed in R with *meta* (for the main meta-analysis) and *metafor* (for meta-regression).

Results

Descriptive Statistics

The demographic information of the 19 included studies is presented in Table 1. All the studies were published after 2023, although the generative pre-trained transformer was announced in 2017. Most of the selected studies were conducted in Asian countries or regions. Fifteen out of 19 studies were done in higher education. GenAI interventions have been widely implemented in various disciplines, including medicine, mathematics, law, and language learning. Regarding the type of GenAI, 14 studies directly adopted general GenAI tools, such as OpenAI's ChatGPT, while the rest developed their own platforms for a specific course based on GenAI technology. Only five studies (26.3%) used GenAI for over a month. In five studies, teachers provided support in the GenAI use process, while in the other 11 studies, no teacher support was provided when students used GenAI for learning purposes. As for the design of the control group, researchers have compared the effects of GenAI with teacher feedback (seven studies), feedback from other resources (six studies), and no feedback (five studies). The quality

Table 1. Study Summary.

Author	Publication year	Country/ region	GenAI tool	Teacher support	Intervention duration	Educational level	Discipline	Control group	Sample size	Study quality
Alneyadi & Wardat (2023)	2023	United Arab Emirates	General	NA	≤ One month	K-12	Science	teacher feedback	122	Weak
Alneyadi & Wardat (2024)	2024	United Arab Emirates	General	NA	NA	K-12	Science	teacher feedback	112	Weak
Baba et al. (2024)	2024	Morocco	Course-specific	With support	≤ One month	Tertiary	Science	feedback from other resources	101	Moderate
Bachiri et al. (2023)	2023	NA	Course-specific	Without support	≤ One month	K-12	NA	no feedback	100	Weak
Dasari et al. (2024)	2024	Indonesia	General	Without support	NA	Tertiary	Science	teacher feedback	29	Moderate
Escalante et al. (2023)	2023	Asia–Pacific region	General	Without support	> One month	Tertiary	Social science	teacher feedback	48	Weak
Hsu (2024)	2024	Taiwan	General	NA	> One month	Tertiary	Science	feedback from other resources	60	Moderate
Li (2023)	2023	Mainland China	Course-specific	With support	≤ One month	Tertiary	Science	teacher feedback	81	Strong
Mahapatra (2024)	2024	India	General	Without support	≤ One month	Tertiary	Science	no feedback	72	Weak
Meyer et al. (2024)	2024	German	General	Without support	≤ One month	K-12	Social science	no feedback	459	Strong

(continued)

Table 1. (continued)

Author	Publication year	Country/ region	GenAI tool	Teacher support	Intervention duration	Educational level	Discipline	Control group	Sample size	Study quality
Niloy et al. (2024)	2024	Bangladesh	General	Without support	≤ One month	Tertiary	Social science	feedback from other resources	600	Moderate
Saravia-Rojas et al. (2024)	2024	Peru	General	Without support	≤ One month	Tertiary	Science	feedback from other resources	55	Moderate
Shi et al. (2024)	2024	China	Course-specific	Without support	≤ One month	Tertiary	Social science	feedback from other resources	128	Weak
Shoufan (2023)	2023	United Arab Emirates	General	Without support	≤ One month	Tertiary	Science	teacher feedback	80–117	Moderate
Sun et al. (2024)	2024	China	Course-specific	Without support	> One month	Tertiary	Science	feedback from other resources	82	Weak
Uddin et al. (2023)	2023	United States	General	With support	≤ One month	Tertiary	Science	NA	42	Strong
Wiboolyasarin et al. (2024)	2024	Thailand	General	Without support	> One month	Tertiary	Social science	no feedback	39	Moderate
Wu et al. (2024)	2024	Mainland China	General	With support	≤ One month	Tertiary	Science	teacher feedback	61	Weak
Zhou and Kim (2024)	2024	Korea	General	With support	> One month	Tertiary	Social science	no feedback	74	Weak

assessment revealed that only three studies have strong quality, with seven medium-quality studies and nine weak-quality studies.

The Overall Effects of GenAI Interventions

Overall, 19 studies with 24 effect sizes reported the comparative effectiveness of GenAI by conducting either a quasi-/experimental study ($n = 17$, $k = 22$) or a repeated-measure study ($n = 2$, $k = 2$). The forest plot (Figure 2) shows that 14 studies revealed a positive effect size, ranging from 0.08 to 4.39. Five studies reported a negative effect size, ranging from −1.86 to −0.18. No outliers were detected in these studies (Hedge's $g > 4.47$ or $g < -3.03$). The overall effect of GenAI interventions on student academic performance was 0.683, significantly different from zero (95% CI: 0.17–1.19; $t = 2.76$, $p = .01 < 0.05$, $k = 24$ in 19 studies). The between-study heterogeneity variance was estimated at $\tau^2 = 1.27$ (95%CI: 0.75–2.88) with $I^2 = 94.4\%$ (95%CI: 92.8%–95.7%), indicating substantial inconsistency between studies. The following moderator analyses reported "teacher support" as a significant moderator. The prediction interval ranged from $g = -1.71$ to 3.08, indicating possible negative intervention effects in future studies. The ANOVA test of the model comparison between the two-level and three-level models (Table 2) showed that the two-level model had a better model fit with lower Akaike (AIC) and Bayesian Information Criterion (BIC). The likelihood ratio test (LRT) result is also not statistically significant ($X^2 = 1.13$, $p = .29$), indicating that the three-level model is unsuitable for the current review. By choosing the two-level model, we ignored independence because only a very small number of studies included more than one effect size; thus, the result of this meta-analysis may not be substantially influenced by treating these effect sizes as independent (van den Noortgate et al., 2013).

Factors Moderating the Effect Size of GenAI on Student Academic Performance

Three groups of moderators were examined in this synthesis: *the implementation of GenAI tools, the context of the study,* and *research design*. The results of the moderator analysis are presented in Table 3. Only one moderator, *teacher support*, showed a significant moderating effect. GenAI interventions showed a significantly larger effect size when teachers were involved in the GenAI process ($g = 1.426$, $p < .01$) than when teachers were missing in the student-GenAI interaction ($g = 0.077$, $p > .05$).

No other significant moderators were identified, but there are some observable differences in the effect sizes of different categories. For example, GenAI used in natural and applied science courses ($g = 0.958$, $p < .01$) had a larger effect size than that used in humanities and social sciences courses ($g = 0.017$, $p > .05$). The mean effect size was larger when using course-specific GenAI tools ($g = 0.848$, $p < .1$) than when using existing GenAI tools ($g = 0.600$, $p < .1$). When GenAI was used in short time durations, the mean effect size ($g = 0.833$, $p < .01$) was larger than when it was used for more than one month ($g = 0.609$, $p > .05$). In terms of research design, the mean effect size of

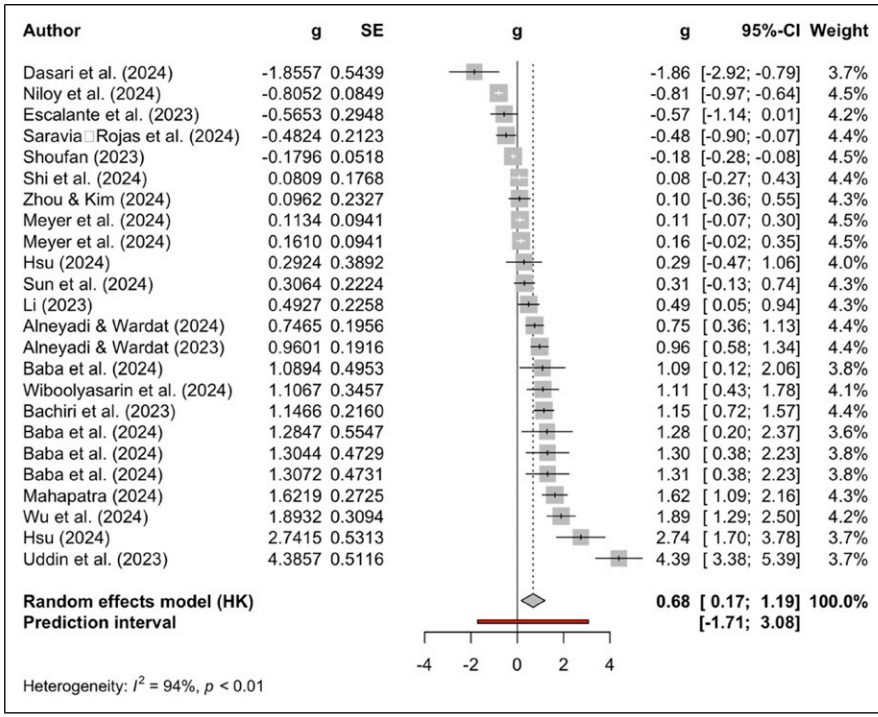

Figure 2. Forest Plots of Effect Sizes.

Table 2. Model Comparison.

	df	AIC	BIC	LRT	*p*
Full	3	79.048	82.455		
Reduced	2	78.180	80.451	1.132	0.287

studies with teacher feedback groups ($g = 0.279$, $p > .05$) was smaller than those with no feedback groups ($g = 0.687$, $p > .05$) and other resources groups ($g = 0.601$, $p > .05$). The mean effect size was bigger in studies with small samples ($g = 0.804$, $p < .05$) than with large samples ($g = 0.560$, $p > .05$). In terms of study quality, GenAI has a larger effect size in studies with high quality ($g = 1.150$, $p > .05$) than in studies with moderate quality ($g = 0.482$, $p > .05$) or weak quality ($g = 0.695$, $p > .05$). However, none of these comparisons revealed a statistically significant difference.

Table 3. Differences in Effect Sizes for Moderators.

Moderators		No. of effect size	Estimate (95%CI)	Test statistic	*p*-value
Implementation of GenAI					
GenAI tool	General GenAI tools	16	0.600 [-0.034; 1.234]	$F(1, 22) = 0.213$	0.649
	Course-specific GenAI tools	8	0.848 [-0.065; 1.760]		
Teacher support	With support	8	1.426 [0.609; 2.242]	$F(1, 18) = 7.565$	0.013*
	Without support	12	0.077 [-0.551; 0.705]		
Intervention duration	≤ One month	16	0.833 [0.225; 1.441]	$F(1, 20) = 0.160$	0.693
	> One month	6	0.609 [-0.394; 1.611]		
Context of the study					
Educational level	K-12	5	0.620 [-0.491; 1.731]	$F(1, 22) = 0.017$	0.899
	Higher education	19	0.698 [0.105; 1.291]		
Discipline	Natural and applied science	16	0.958 [0.342; 1.574]	$F(1, 21) = 3.256$	0.086
	Humanities and social science	7	0.017 [-0.876; 0.909]		
Research design					
Control group	No feedback	6	0.687 [-0.128; 1.502]	$F(2, 20) = 0.337$	0.718
	Teacher feedback	7	0.279 [-0.490; 1.047]		
	Feedback from other resources	10	0.601 [-0.068; 1.270]		
Sample size	Small (<100)	12	0.804 [0.063; 1.545]	$F(1, 22) = 0.236$	0.632
	Large (≥100)	12	0.560 [-0.171; 1.291]		
Study quality	Strong	4	1.150 [-0.121; 2.422]	$F(2, 21) = 0.430$	0.656
	Moderate	11	0.482 [-0.312; 1.276]		
	Weak	9	0.695 [-0.146; 1.537]		

Note. *$p < .05$. *Summary*: • Significant variation was revealed between studies with and without teacher support. • Other characteristics did not show significant variations.

Sensitivity Analyses

A sensitivity analysis was performed using the leave-one-out method, sorted by the pooled effect size (See Figure 3). The results show that the original pooled effect size will not be influenced when leaving out each study, as the changed effect sizes still fall within the 95% confidence interval of the original pooled effect size (0.17–1.19).

A sensitivity analysis excluding studies with weak qualities was also conducted, and the results are presented in Table 4. The overall effects increased slightly after removing weak studies (from g = 0.683–0.692).

Publication Bias

The funnel plot (See Figure 4) and the statistical data from Egger's regression test (β = 4.03, t = 3.65, p = .014) showed that the data were asymmetrical. However, Vevea and Woods' (2005) selection model revealed minimal adjustment from 0.683 to 0.681. This result showed that the observed asymmetry may not be caused by publication bias.

Discussion

The relatively new AI technology, GenAI, has the potential to facilitate student learning by providing timely and personalised feedback (Stojanov, 2023); however, overreliance on GenAI tools could exert harmful effects (Susnjak & McIntosh, 2024). This meta-analysis is among the first few attempts to investigate exclusively the effects of GenAI interventions on student academic performance. After examining 24 effect sizes from 19 empirical studies across various disciplines with either experimental-control design or pre-post comparison design, key meta-analytic results were presented as follows:

(1) An overall large effect size (g = 0.683) was found on the effects of GenAI interventions on student academic performance.
(2) GenAI interventions had a more pronounced effect on students receiving teacher support (g = 1.426) than on those without teacher support in the student-GenAI interaction (g = 0.077).
(3) No statistically significant differences were found in effect sizes across different GenAI tools, intervention duration, educational level, discipline, control group, sample size, and study quality.

Overall Effects of GenAI Interventions

The mean effect size of the GenAI interventions in this synthesis is 0.683, suggesting a large effect size in educational interventions (Hattie, 2008). This result supports the claims that GenAI has the potential to facilitate student learning by working as more knowledgeable others and providing learners with personalised scaffolding, real-time

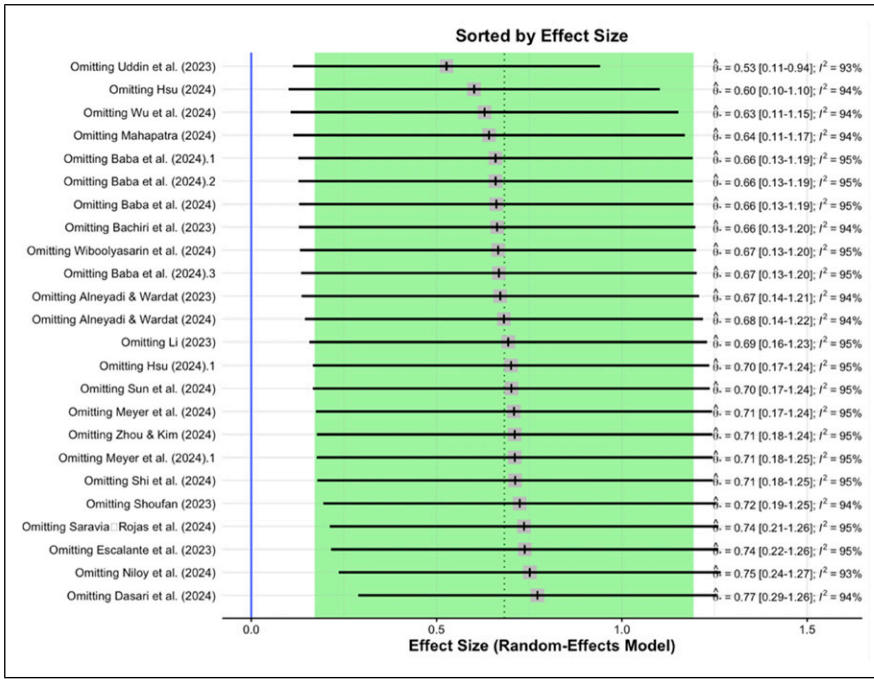

Figure 3. Sensitivity Analysis Leaving Out Each Study.

Table 4. Sensitivity Analysis Excluding Weak Studies.

	g	95%CI	*p*	95%PI	I^2	95%CI
Meta analysis	0.683	0.172–1.194	0.011	−1.71–3.08	94.4%	92.8–95.7
Weak studies Removed*	0.692	−0.115-1.50	0.087	−2.40–3.78	94.5%	92.3–96.0

Note. *Removed as weak studies: Alneyadi & Wardat (2023), Alneyadi & Wardat (2024), Bachiri et al. (2023), Escalante et al. (2023), Mahapatra (2024), Shi et al. (2024), Sun et al. (2024), Wu et al. (2024), Zhou and Kim (2024).

feedback, and interactions (Darvishi et al., 2024; Stojanov, 2023). Students may also have more positive psychological reactions (e.g., higher motivation, Song & Song, 2023; higher self-efficacy, Teng, 2024; and less nervousness, Hsu et al., 2023) when interacting with GenAI tools.

The mean effect size of the current study was larger than that revealed in Sun and Zhou's (2024) meta-analysis ($g = 0.533$). The higher effect size may be attributed to the broader educational scope of the current review, including both K-12 and college students, while Sun and Zhou (2024) only focused on higher education. Zheng et al.

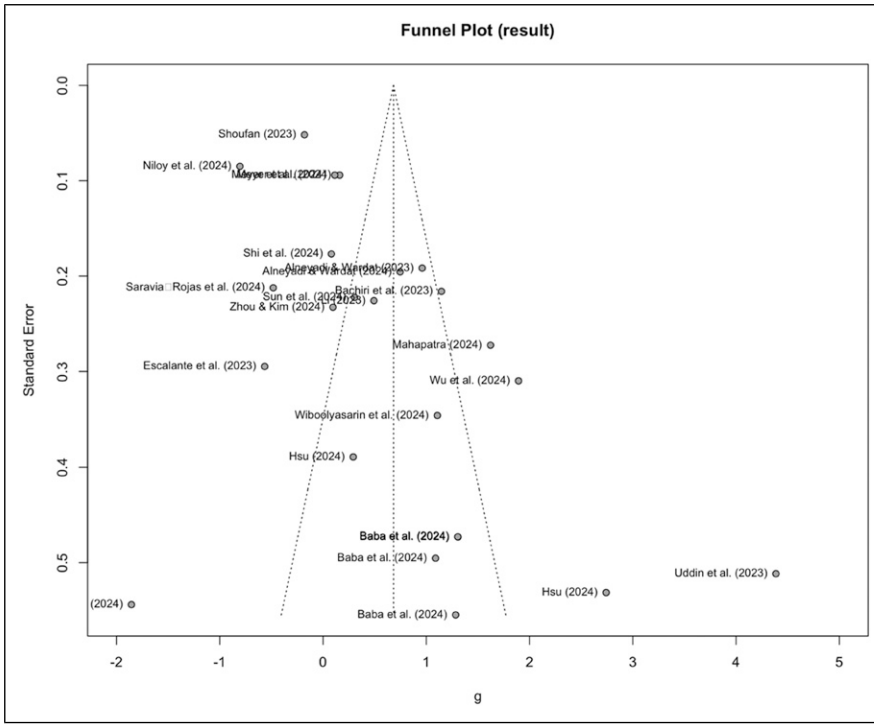

Figure 4. Funnel Plot.

(2023) found that high school students benefited more than post-secondary students; hence, including K-12 students is likely to increase the mean effect size. Although our findings showed that the educational level did not significantly moderate the effects of GenAI interventions on student academic achievement, such a result needs to be interpreted with caution because of the limited number of primary studies in the K-12 context. This synthesis also showed a larger mean effect size than that in Wang et al.'s (2024a) meta-analysis ($g = 0.484$). This is probably because Wang et al. (2024a) only considered the language learning domain, while this synthesis included studies across different disciplines. Compared with the effect size in the meta-analyses of Wu and Yu (2024) ($d = 1.028$) and Zheng et al. (2023) ($g = 0.812$), which investigated AI technology in general, this study revealed a smaller mean effect size. GenAI is a relatively new technology under the umbrella of AI technology. The related innovations may not have been well implemented and validated in naturalistic educational settings (Yan et al., 2024). Researchers, teachers, and students need time to explore and adjust its optimal use in educational practice.

Despite the overall positive effects of GenAI interventions on student academic performance, five out of 19 studies (26.32%) reported a negative effect size, showing

that the positive effects of GenAI interventions are not warranted, and the use of this tool needs careful design and implementation. Across all five studies, teachers were not involved in students' GenAI use, suggesting that this lack of teacher support may have contributed to the negative effects. While GenAI can help automate some educational tasks (e.g., providing feedback and generating questions), teachers' emotional support, moral guidance, and expertise in specific domains can all be valuable in scaffolding students in the process of GenAI interventions (Tam, 2024). Moreover, these five studies had relatively short intervention durations: three used GenAI tools for less than one month, one had a six-week intervention, and one did not provide relevant information. This result indicates that short GenAI interventions will likely generate negative learning gains, contradicting Wu and Yu's (2024) finding that students gained more learning benefits in shorter interventions (less than ten weeks). Such a difference is probably because of the definition of the short duration of the intervention (i.e., one month or ten weeks as the threshold) and students' less familiarity with GenAI tools than general AI chatbots. Although all five studies were conducted in the tertiary context using general GenAI tools, these findings warrant cautious interpretation because of the limited number of comparative studies - more than ten studies each in higher education and using general GenAI tools, but only four studies in K-12 education and six studies using course-specific GenAI tools.

Moderators of GenAI Interventions

Three groups of moderators were examined in this meta-analysis: implementation of GenAI, context of the study, and research design. A key contribution of our study is to identify "teacher support" that could significantly moderate the effects of GenAI interventions, which was not considered in the study of Sun and Zhou (2024). While Sun and Zhou (2024) explored the moderators from the perspective of designable pedagogy, teachers also play an important role in students' learning using GenAI tools. We found that students could gain more learning benefits after using GenAI tools with teacher support than those who solely rely on GenAI in the learning process. This finding supported the claim that teachers should provide students with scaffolding and supplementary feedback so that students could better incorporate GenAI feedback into their work; otherwise, students may struggle to critically evaluate the GenAI outputs (Su et al., 2023). Han and Li (2024) also emphasised the role of teachers by proposing an "AI + Teacher" model, which argued for making the best use of both the analytical strengths of AI and the pedagogical expertise of instructors. In this way, students can optimise the use of GenAI outputs in their study while maintaining teacher-student interactions. The importance of teachers' roles has not been diminished; instead, teachers' pedagogical decisions are vital regarding using GenAI tools (Jeon & Lee, 2023).

Apart from teacher support, no other significant moderators were detected in this synthesis, although the importance of these hypothesised moderators was theoretically supported. These results are inconsistent with previous reviews that identified other significant moderators (e.g., educational level and intervention duration, Wu & Yu,

2024; sample size, Sun & Zhou, 2024; Zheng et al., 2023). One reason may be that the number of included studies and effect sizes in some categories is too small to detect significant moderators. Fu et al. (2011) suggested that at least four studies in each subgroup be the lower bound for categorical moderator analysis. Therefore, in the "educational level" dimension, due to limited evidence from primary education, this study combined K-12 studies rather than analysing GenAI interventions separately for primary and secondary students. This aggregation may have obscured intervention effects, as primary and secondary students differ in language proficiency, digital literacy, self-regulated learning abilities, and academic pressures (Jeon, 2024; Tang et al., 2020). In addition, the nonsignificant moderator effect of "intervention duration" may stem from using "one month" as the dividing threshold. While Lo et al. (2024) recommended a whole semester of implementation to mitigate the novelty effect of ChatGPT, few studies in this synthesis maintained such long durations. The one-month threshold may be insufficient to detect significant differences, as students may still get familiar with this new technology over a five- or six-week implementation. Another reason could be the collinearity of moderators (Murano et al., 2020). For instance, one study sampled university students who had a two-month intervention (Hsu, 2024), while one study selected K-12 students and implemented only a 90-minute session (Meyer et al., 2024). Such a diversity of the characteristics of selected studies may cause inaccurate estimates of individual moderators.

Implications

Researchers argued that GenAI could promote student learning by providing students with timely feedback from various perspectives (Xia et al., 2024) and facilitating student self-directed learning (Yu, 2024). This meta-analysis supports the argument that GenAI could positively affect student academic achievement across contexts. Learners can use GenAI tools for various purposes, such as a virtual intelligent assistant to get instant feedback, a writing assistant to enhance writing skills, or an aiding tool to gain a personalised learning experience (Albadarin et al., 2024). Hence, rather than restricting the use of GenAI, educational institutions should implement guidelines encouraging its integration into teaching and learning processes. These policies should emphasise the distinct value of human instructors and the limitations of GenAI tools (An et al., 2025). Additionally, institutional policies should support students in developing prompt engineering skills to gain high-quality outputs, avoiding the misuse of GenAI tools (Knoth et al., 2024).

Despite the overall positive effects of GenAI on student academic performance, negative impacts were also observed in some cases, as indicated in this meta-analysis (26.32% of the included studies). Hence, careful design and implementation are needed to avoid the harmful effects. Considering that the "teacher support" factor significantly moderated the GenAI effects, teachers are encouraged to proactively participate in students' dialogue with GenAI tools. Previous research showed that teachers could support students by specifying learning objects before the integration of GenAI tools

(Su & Yang, 2023), enhancing students' prompting strategies and meta-cognitive skills (Zhan & Yan, 2025), demonstrating the use of GenAI tools and discussing with students the ethical issues of using GenAI tools (Moorhouse et al., 2024). This meta-analysis enriched current understanding by identifying additional effective strategies from the selected studies, including discussing GenAI-suggested content with students (Wu et al., 2024), providing ongoing support during students' interactions with GenAI instead of merely the training before the intervention (Uddin et al., 2023), identifying and explaining GenAI feedback errors (Zhou & Kim, 2024), and teachers being part of the GenAI-based learning system who can upload the teaching materials (Baba et al., 2024), design learning sheet and assessments, and also monitor the learning process (Li, 2023).

Since teachers are not born with the capacity to use GenAI tools properly, professional training is needed to equip teachers with the skills to integrate GenAI tools into classroom activities according to instructional goals (Liu & Xiao, 2024). The effective teacher support strategies identified in this meta-analysis align with Kong et al.'s (2024) teacher professional development framework, which emphasises two components: developing teachers' AI literacy and fostering their ability to implement student-centred pedagogy when incorporating GenAI in teaching. Teachers need to develop AI literacy to understand GenAI's capabilities and limitations (e.g., errors in responses and prompt engineering techniques) and become proficient with GenAI teaching tools (e.g., GenAI-based learning platforms). Moreover, with student-centred pedagogical skills, teachers could know how to guide students in effective AI use through modelling and discussion. While only one significant moderator is detected in this synthesis, teachers are still suggested to consider and reflect on when and how to use GenAI tools based on students' needs to optimise the positive influence on student academic achievement.

Findings from this study can also inform the development of GenAI-based educational tools. While most of the current GenAI tools are not primarily designed for education, as found in most of the included studies (14 out of 19, 73.68%), their effective implementation requires consistent teacher support and guidance. Hence, curriculum designers should strategically align GenAI technology and effective teaching strategies with curriculum standards, classroom content, and teaching objectives (Wang et al., 2024b). As GenAI development and deployment become more cost-effective (Ferrara, 2024), there is also an opportunity for researchers or technicians to develop course-specific GenAI systems. These systems should incorporate teacher roles to facilitate meaningful teacher involvement, such as the platform in Baba et al.'s (2024) study. Such embedded teacher support allows teachers to effectively guide and support student learning while maximising the benefits of GenAI technology.

Limitations and Future Studies

The current meta-analysis has several limitations. First, only 19 studies, including a very small number of high-quality studies (15.78%), were included in this synthesis.

Although the sensitivity analysis excluding weak studies showed a slight difference, such a small sample size with quite a high proportion of weak studies may limit generalizability. Considering the rapid emergence of empirical studies about GenAI interventions, future reviews can include more eligible and high-quality studies to update our understanding in this area. Second, this study only focuses on the effects of GenAI interventions on student academic performance. Apart from cognitive outcomes, GenAI may also influence self-regulated learning (Lee et al., 2024) and social-emotional outcomes (e.g., motivation and attitudes toward GenAI) (Salas-Pilco, 2020). Future studies could investigate non-academic learning outcomes, such as motivation, self-efficacy, self-regulated learning, and higher-order thinking skills. Third, because of the missing information and limited variability in this group of studies, this synthesis did not investigate several personal and contextual factors that might contribute to the high heterogeneity, potentially limiting the generalizability of the findings. For instance, training students before using GenAI tools (Abdelhalim, 2024), along with their familiarity (Wood & Moss, 2024) and self-efficacy (Tantivejakul et al., 2024) in using these technologies, can influence their acceptance and effective use of this novel technology. Pedagogical methods in which GenAI is integrated can also vary across different studies (e.g., ChatGPT-based flipped learning, Li, 2023; using ChatGPT for self- and peer assessment, Mahapatra, 2024). Future research could explore these important factors that may influence the effects of GenAI interventions. Additionally, we imputed a correlation of 0.5 when using Becker's (1988) formula to calculate the effect size of Cohen's *d* for studies with repeated-measures design. Instead of conducting a sensitivity analysis of a wide range of correlations, using a fixed value of correlation to estimate the effect size may bias the true treatment effect because of not capturing comprehensive possibilities in the selected studies (Cuijpers et al., 2017).

Conclusion

This meta-analysis investigated the effects of GenAI interventions on student academic performance. The results showed that overall, GenAI interventions positively affected student academic performance. This finding supported the theoretical arguments that GenAI has the potential to promote student learning. However, negative effects were observed in some studies, encouraging teachers and students to implement this novel AI technology in practice with careful design. This meta-analysis also revealed that students with teacher support in GenAI interventions gained significantly larger learning benefits in the use of GenAI tools than students solely dependent on GenAI tools, suggesting an indispensable role of teachers in students' interaction with GenAI tools. Alongside the rapid development of GenAI technology, more studies are needed to scrutinise how to use GenAI tools effectively and optimise their impact on learning.

Acknowledgements

The authors acknowledge and thank Peiyao Zhang's careful screening and coding.

Declaration of Conflicting Interests

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

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Ethical Statement

Ethical Approval

No information was directly collected from human participants.

ORCID iDs

Jiahe Gu https://orcid.org/0000-0002-5757-8191
Zi Yan https://orcid.org/0000-0001-9305-884X

Data Availability Statement

Data will be available on request.

Supplemental Material

Supplemental material for this article is available online.

References

References marked with an asterisk indicate studies included in the meta-analysis.

Abdelhalim, S. M. (2024). Using ChatGPT to promote research competency: English as a foreign language undergraduates' perceptions and practices across varied metacognitive awareness levels. *Journal of Computer Assisted Learning*, *40*(3), 1261–1275. https://doi.org/10.1111/jcal.12948

Acuna, E., & Rodriguez, C. (2004). *A meta-analysis study of outlier detection methods in classification*. Department of Mathematics, University of Puerto Rico at Mayaguez. https://www.researchgate.net/publication/228728761_A_meta_analysis_study_of_outlier_detection_methods_in_classification

Adeshola, I., & Adepoju, A. P. (2023). The opportunities and challenges of ChatGPT in education. *Interactive Learning Environments*, 1–14. https://doi.org/10.1080/10494820.2023.2253858

Albadarin, Y., Saqr, M., Pope, N., & Tukiainen, M. (2024). A systematic literature review of empirical research on ChatGPT in education. *Discover Education*, *3*(1), 60. https://doi.org/10.1007/s44217-024-00138-2

Almasre, M. (2024). Development and evaluation of a custom GPT for the assessment of students' designs in a typography course. *Education Sciences (Basel)*, *14*(2), 148. https://doi.org/10.3390/educsci14020148

Alshraah, S. M., Kariem, N., Alshraah, A. M., Aldosemani, T. I., & AlQarni, K. (2024). A critical look at how lecturers in linguistics can leverage generative artificial intelligence in enhancing teaching proficiency and students' engagement. *Journal of Language Teaching and Research*, *15*(4), 1361–1371. https://doi.org/10.17507/jltr.1504.34

An, Y., Yu, J. H., & James, S. (2025). Investigating the higher education institutions' guidelines and policies regarding the use of generative AI in teaching, learning, research, and administration. *International Journal of Educational Technology in Higher Education*, *22*(1), 10. https://doi.org/10.1186/s41239-025-00507-3

Ba, H., Zhang, L., & Yi, Z. (2024). Enhancing clinical skills in pediatric trainees: A comparative study of ChatGPT-assisted and traditional teaching methods. *BMC Medical Education*, *24*(1), 558. https://doi.org/10.1186/s12909-024-05565-1

* Baba, K., Faddouli, E., & Cheimanoff, N. (2024). Mobile-optimised AI-driven personalised learning: A case study at Mohammed VI polytechnic university. *International Journal of Interactive Mobile Technologies*, *18*(4). https://doi.org/10.3991/ijim.v18i04.46547

Baber, H., Nair, K., Gupta, R., & Gurjar, K. (2023). The beginning of ChatGPT–a systematic and bibliometric review of the literature. *Information and Learning Sciences*, *125*(7/8), 587–614. https://doi.org/10.1108/ILS-04-2023-0035

* Bachiri, Y. A., Mouncif, H., & Bouikhalene, B. (2023). Artificial intelligence empowers gamification: Optimising student engagement and learning outcomes in e-learning and moocs. *International Journal of Engineering Pedagogy*, *13*(8), 1. https://doi.org/10.3991/ijep.v13i8.40853

Bahroun, Z., Anane, C., Ahmed, V., & Zacca, A. (2023). Transforming education: A comprehensive review of generative artificial intelligence in educational settings through bibliometric and content analysis. *Sustainability (Basel)*, *15*(17), 12983. https://doi.org/10.3390/su151712983

Banihashem, S. K., Kerman, N. T., Noroozi, O., Moon, J., & Drachsler, H. (2024). Feedback sources in essay writing: Peer-generated or AI-generated feedback? *International Journal of Educational Technology in Higher Education*, *21*(1), 23. https://doi.org/10.1186/s41239-024-00455-4

Barrett, A., & Pack, A. (2023). Not quite eye to AI: Student and teacher perspectives on the use of generative artificial intelligence in the writing process. *International Journal of Educational Technology in Higher Education*, *20*(1), 59. https://doi.org/10.1186/s41239-023-00427-0

Bearman, M., Tai, J., Dawson, P., Boud, D., & Ajjawi, R. (2024). Developing evaluative judgement for a time of generative artificial intelligence. *Assessment & Evaluation in Higher Education*, *49*(6), 893–905. https://doi.org/10.1080/02602938.2024.2335321

Becker, B. J. (1988). Synthesising standardised mean-change measures. *British Journal of Mathematical and Statistical Psychology*, *41*(2), 257–278. https://doi.org/10.1111/j.2044-8317.1988.tb00901.x

Bengesi, S., El-Sayed, H., Sarker, M. K., Houkpati, Y., Irungu, J., & Oladunni, T. (2024). Advancements in generative AI: A comprehensive review of GANs, GPT, autoencoders,

diffusion model, and transformers. *IEEE Access*, *12*, 69812–69837. https://doi.org/10.1109/ACCESS.2024.3397775

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2021). *Introduction to meta-analysis*. John wiley & sons.

Bower, M., Torrington, J., Lai, J. W., Petocz, P., & Alfano, M. (2024). How should we change teaching and assessment in response to increasingly powerful generative artificial intelligence? Outcomes of the ChatGPT teacher survey. *Education and Information Technologies*, *29*, 15403–15439. https://doi.org/10.1007/s10639-023-12405-0

Cheah, Y. H., & Kim, J. (2025). STEM teachers' perceptions, familiarity, and support needs for integrating generative artificial intelligence in K-12 education. *School Science & Mathematics*, 1–16. https://doi.org/10.1111/ssm.18334

Chiu, T. K. (2024). The impact of generative AI (GenAI) on practices, policies and research direction in education: A case of ChatGPT and Midjourney. *Interactive Learning Environments*, *32*(10), 6187–6203. https://doi.org/10.1080/10494820.2023.2253861

Coenen, C., & Pfenninger, M. (2024). Transforming learning experiences and assessments through AI-empowered cocreation of quality feedback. *New Directions for Teaching and Learning*. https://doi.org/10.1002/tl.20628

Cohen, J. C. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Erlbaum Associates.

Crawford, J., Vallis, C., Yang, J., Fitzgerald, R., O'dea, C., & Cowling, M. (2023). Artificial intelligence is awesome, but good teaching should always come first. *Journal of University Teaching and Learning Practice*, *20*(7), 01. https://doi.org/10.53761/1.20.7.01

Cuijpers, P., Weitz, E., Cristea, I. A., & Twisk, J. (2017). Pre-post effect sizes should be avoided in meta-analyses. *Epidemiology and Psychiatric Sciences*, *26*(4), 364–368. https://doi.org/10.1017/S2045796016000809

Darvishi, A., Khosravi, H., Sadiq, S., Gašević, D., & Siemens, G. (2024). Impact of AI assistance on student agency. *Computers & Education*, *210*, 104967. https://doi.org/10.1016/j.compedu.2023.104967

* Dasari, D., Hendriyanto, A., Sahara, S., Suryadi, D., Muhaimin, L. H., Chao, T., & Fitriana, L. (2024). ChatGPT in didactical tetrahedron, does it make an exception? A case study in mathematics teaching and learning. *Frontiers in Education*, *8*, 1295413. https://doi.org/10.3389/feduc.2023.1295413

Egger, M., Davey Smith, G. D., 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

* Escalante, J., Pack, A., & Barrett, A. (2023). AI-generated feedback on writing: Insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education*, *20*(1), 57. https://doi.org/10.1186/s41239-023-00425-2

Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, D., Li, X., & Gašević, D. (2025). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology*, *56*(2), 489–530. https://doi.org/10.1111/bjet.13544

Ferrara, E. (2024). GenAI against humanity: Nefarious applications of generative artificial intelligence and large language models. *Journal of Computational Social Science*, *7*(1), 549–569. https://doi.org/10.1007/s42001-024-00250-1

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, *76*(5), 378–382. https://doi.org/10.1037/h0031619

Fu, R., Gartlehner, G., Grant, M., Shamliyan, T., Sedrakyan, A., Wilt, T. J., Griffith, T. A., Oremus, M., Raina, P., Ismaila, A., Santaguida, P., Lau, J., & Trikalinos, T. A. (2011). Conducting quantitative synthesis when comparing medical interventions: AHRQ and the effective health care program. *Journal of Clinical Epidemiology*, *64*(11), 1187–1197. https://doi.org/10.1016/j.jclinepi.2010.08.010

Ghafouri, M., Hassaskhah, J., & Mahdavi-Zafarghandi, A. (2024). From virtual assistant to writing mentor: Exploring the impact of a ChatGPT-based writing instruction protocol on EFL teachers' self-efficacy and learners' writing skill. *Language Teaching Research*, 13621688241239764. https://doi.org/10.1177/13621688241239764

Google. (2023). What is generative AI and what are its applications? Google cloud. https://cloud.google.com/use-cases/generative-ai

Guo, Y., & Lee, D. (2023). Leveraging ChatGPT for enhancing critical thinking skills. *Journal of Chemical Education*, *100*(12), 4876–4883. https://doi.org/10.1021/acs.jchemed.3c00505

Habib, S., Vogel, T., Anli, X., & Thorne, E. (2024). How does generative artificial intelligence impact student creativity? *Journal of Creativity*, *34*(1), 100072. https://doi.org/10.1016/j.yjoc.2023.100072

Han, J., & Li, M. (2024). Exploring ChatGPT-supported teacher feedback in the EFL context. *System (Linköping)*, *126*, 103502. https://doi.org/10.1016/j.system.2024.103502

Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, *6*(2), 107–128. https://doi.org/10.3102/10769986006002107

Hsu, H. L., Chen, H. H. J., & Todd, A. G. (2023). Investigating the impact of the Amazon Alexa on the development of L2 listening and speaking skills. *Interactive Learning Environments*, *31*(9), 5732–5745. https://doi.org/10.1080/10494820.2021.2016864

* Hsu, M. H. (2024). Mastering medical terminology with ChatGPT and Termbot. *Health Education Journal*, *83*(4), 352–358. https://doi.org/10.1177/00178969231197371

Hwang, S. (2022). Examining the effects of artificial intelligence on elementy students' mathematics achievement: A meta-analysis. *Sustainability (Basel)*, *14*(20), 13185. https://doi.org/10.3390/su142013185

Imran, M., & Almusharraf, N. (2023). Analysing the role of ChatGPT as a writing assistant at higher education level: A systematic review of the literature. *Contemporary Educational Technology*, *15*(4), ep464. https://doi.org/10.30935/cedtech/13605

Jeon, J. (2024). Exploring AI chatbot affordances in the EFL classroom: Young learners' experiences and perspectives. *Computer Assisted Language Learning*, *37*(1-2), 1–26. https://doi.org/10.1080/09588221.2021.2021241

Jeon, J., & Lee, S. (2023). Large language models in education: A focus on the complementary relationship between human teachers and ChatGPT. *Education and Information Technologies*, *28*(12), 15873–15892. https://doi.org/10.1007/s10639-023-11834-1

Kalota, F. (2024). A primer on generative artificial intelligence. *Education Sciences (Basel)*, *14*(2), 172. https://doi.org/10.3390/educsci14020172

Keath, A., Wyant, J., & Towner, B. (2024). ChatGPE: Does artificial intelligence have a place in the physical education setting? *Journal of Physical Education, Recreation and Dance*, *95*(2), 59–61. https://doi.org/10.1080/07303084.2023.2292941

Kelly, A., Sullivan, M., & Strampel, K. (2023). Generative artificial intelligence: University student awareness, experience, and confidence in use across disciplines. *Journal of University Teaching and Learning Practice*, *20*(6), 12. https://doi.org/10.53761/1.20.6.12

Kietzmann, J., & Park, A. (2024). Written by ChatGPT: AI, large language models, conversational chatbots, and their place in society and business. *Business Horizons*. https://doi.org/10.1016/j.bushor.2024.06.002

Kizilcec, R. F., Huber, E., Papanastasiou, E. C., Cram, A., Makridis, C. A., Smolansky, A., Zeivots, C., & Raduescu, C. (2024). Perceived impact of generative AI on assessments: Comparing educator and student perspectives in Australia, Cyprus, and the United States. *Computers and Education: Artificial Intelligence*, *7*, 100269. https://doi.org/10.1016/j.caeai.2024.100269

Knoth, N., Tolzin, A., Janson, A., & Leimeister, J. M. (2024). AI literacy and its implications for prompt engineering strategies. *Computers and Education: Artificial Intelligence*, *6*, 100225. https://doi.org/10.1016/j.caeai.2024.100225

Kong, S. C., Yang, Y., & Hou, C. (2024). Examining teachers' behavioural intention of using generative artificial intelligence tools for teaching and learning based on the extended technology acceptance model. *Computers and Education: Artificial Intelligence*, *7*, 100328. https://doi.org/10.1016/j.caeai.2024.100328

Law, L. (2024). Application of generative artificial intelligence (GenAI) in language teaching and learning: A scoping literature review. *Computers and Education Open*, *6*, 100174. https://doi.org/10.1016/j.caeo.2024.100174

Lee, H., & Lee, J. H. (2024). The effects of AI-guided individualised language learning: A meta-analysis. *Language, Learning and Technology*, *28*(2), 134–162. https://hdl.handle.net/10125/73575

Lee, H. Y., Chen, P. H., Wang, W. S., Huang, Y. M., & Wu, T. T. (2024). Empowering ChatGPT with guidance mechanism in blended learning: Effect of self-regulated learning, higher-order thinking skills, and knowledge construction. *International Journal of Educational Technology in Higher Education*, *21*(1), 16. https://doi.org/10.1186/s41239-024-00447-4

Leiter, C., Zhang, R., Chen, Y., Belouadi, J., Larionov, D., Fresen, V., & Eger, S. (2024). ChatGPT: A meta-analysis after 2.5 months. *Machine Learning with Applications*, *16*, 100541. https://doi.org/10.1016/j.mlwa.2024.100541

Li, H. (2023). Effects of a ChatGPT-based flipped learning guiding approach on learners' courseware project performances and perceptions. *Australasian Journal of Educational Technology*, *39*(5), 40–58. https://doi.org/10.14742/ajet.8923

* Li, J., Zong, H., Wu, E., Wu, R., Peng, Z., Zhao, J., Yang, B., Xie, H., & Shen, B. (2024). Exploring the potential of artificial intelligence to enhance the writing of English academic papers by non-native English-speaking medical students-the educational application of ChatGPT. *BMC Medical Education*, *24*(1), 736. https://doi.org/10.1186/s12909-024-05738-y

Li, X., & Xiao, Y. (2024). Chinese university teachers' engagement with generative AI in different stages of foreign language teaching: A qualitative enquiry through the prism of ADDIE. *Education and Information Technologies*, 1–24. https://doi.org/10.1007/s10639-024-13117-9

Liang, H. Y., Hwang, G. J., Hsu, T. Y., & Yeh, J. Y. (2024). Effect of an AI-based chatbot on students' learning performance in alternate reality game-based museum learning. *British Journal of Educational Technology*, *55*(5), 2315–2338. https://doi.org/10.1111/bjet.13448

Liao, X., Zhang, X., Wang, Z., & Luo, H. (2024). Design and implementation of an AI-enabled visual report tool as formative assessment to promote learning achievement and self-regulated learning: An experimental study. *British Journal of Educational Technology*, *55*(3), 1253–1276. https://doi.org/10.1111/bjet.13424

Light, R. J., & Pillemer, D. B. (1984). *Summing up: The science of reviewing research*. Harvard University Press.

Liu, Y., Awang, H., & Mansor, N. S. (2024). To explore Sora's potential influence on future education. *Educational Research*, *6*(5), 98–105.

Lo, C. K., Hew, K. F., & Jong, M. S. Y. (2024). The influence of ChatGPT on student engagement: A systematic review and future research agenda. *Computers & Education*, *219*, 105100. https://doi.org/10.1016/j.compedu.2024.105100

* Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: A mixed methods intervention study. *Smart Learning Environments*, *11*(1), 9. https://doi.org/10.1186/s40561-024-00295-9

Malecki, C. K., & Demaray, M. K. (2002). Measuring perceived social support: Development of the child and adolescent social support scale (CASSS). *Psychology in the Schools*, *39*(1), 1–18. https://doi.org/10.1002/pits.10004

McMillan, J. H., Venable, J. C., & Varier, D. (2013). Studies of the effect of formative assessment on student achievement: So much more is needed. *Practical Assessment, Research and Evaluation*, *18*(2), 1–15. https://doi.org/10.7275/tmwm-7792

* Meyer, J., Jansen, T., Schiller, R., Liebenow, L. W., Steinbach, M., Horbach, A., & Fleckenstein, J. (2024). Using LLMs to bring evidence-based feedback into the classroom: AI-generated feedback increases secondary students' text revision, motivation, and positive emotions. *Computers and Education: Artificial Intelligence*, *6*, 100199. https://doi.org/10.1016/j.caeai.2023.100199

Moorhouse, B. L., Wan, Y., Wu, C., Kohnke, L., Ho, T. Y., & Kwong, T. (2024). Developing language teachers' professional generative AI competence: An intervention study in an initial language teacher education course. *System*, *125*, 103399. https://doi.org/10.1016/j.system.2024.103399

Murano, D., Sawyer, J. E., & Lipnevich, A. A. (2020). A meta-analytic review of preschool social and emotional learning interventions. *Review of Educational Research*, *90*(2), 227–263. https://doi.org/10.3102/0034654320914743

* Niloy, A. C., Akter, S., Sultana, N., Sultana, J., & Rahman, S. I. U. (2024). Is ChatGPT a menace for creative writing ability? An experiment. *Journal of Computer Assisted Learning*, *40*(2), 919–930. https://doi.org/10.1111/jcal.12929

Noetel, M., Griffith, S., Delaney, O., Sanders, T., Parker, P., del Pozo Cruz, B., & Lonsdale, C. (2021). Video improves learning in higher education: A systematic review. *Review of Educational Research*, *91*(2), 204–236. https://doi.org/10.3102/0034654321990713

Olivera-Aguilar, M., Lee, H. S., Pallant, A., Belur, V., Mulholland, M., & Liu, O. L. (2022). Comparing the effect of contextualized versus generic automated feedback on students' scientific argumentation. *ETS Research Report Series*, *2022*(1), 1–14. https://doi.org/10.1002/ets2.12344

OpenAI. (2022). Introducing ChatGPT. https://openai.com/blog/chatgpt

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, D., 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. *International Journal of Surgery*, *88*, 105906. https://doi.org/10.1136/bmj.n71

Panadero, E., & Lipnevich, A. A. (2022). A review of feedback models and typologies: Towards an integrative model of feedback elements. *Educational Research Review*, *35*, 100416. https://doi.org/10.1016/j.edurev.2021.100416

Peñalvo, F. J. G., & Ingelmo, A. V. (2023). What do we mean by GenAI? A systematic mapping of the evolution, trends, and techniques involved in generative AI. *IJIMAI*, *8*(4), 7–16. https://doi.org/10.9781/ijimai.2023.07.006

Preiksaitis, C., & Rose, C. (2023). Opportunities, challenges, and future directions of generative artificial intelligence in medical education: Scoping review. *JMIR Medical Education*, *9*, e48785. https://doi.org/10.2196/48785

Rashidi, H. H., Hu, B., Pantanowitz, J., Tran, N., Liu, S., Chamanzar, A., & Hanna, M. G. (2024). Statistics of generative AI & non-generative predictive analytics machine learning in medicine. *Modern Pathology*, *1*, 100663. https://doi.org/10.1016/j.modpat.2024.100663

Sabzalieva, E., & Valentini, A. (2023). *ChatGPT and artificial intelligence in higher education: Quick start guide*. UNESCO.

Saish, N. V. P., Jayashree, J., & Vijayashree, J. (2025). Mathematical foundations and applications of generative AI models. In N. R. Vajjhala, S. S. Roy, B. Taşci, & M. E. H. Chowdhury (Eds.), *Generative artificial intelligence (AI) approaches for industrial applications* (pp. 19–45). Springer.

Salas-Pilco, S. Z. (2020). The impact of AI and robotics on physical, social-emotional and intellectual learning outcomes: An integrated analytical framework. *British Journal of Educational Technology*, *51*(5), 1808–1825. https://doi.org/10.1111/bjet.12984

Shamseer, L., Moher, D., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, L. A., & Stewart, L. A. (2015). Preferred reporting items for systematic review and meta-analysis

protocols (PRISMA-P) 2015: Elaboration and explanation. *BMJ*, *350*, g7647. https://doi.org/10.1136/bmj.g7647

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, *78*(1), 153–189. https://doi.org/10.3102/0034654307313795

Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: Assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in Psychology*, *14*, 1260843. https://doi.org/10.3389/fpsyg.2023.1260843

Stojanov, A. (2023). Learning with ChatGPT 3.5 as a more knowledgeable other: An autoethnographic study. *International Journal of Educational Technology in Higher Education*, *20*(1), 35. https://doi.org/10.1186/s41239-023-00404-7

Strobel, G., Banh, L., Möller, F., & Schoormann, T. (2024). Exploring generative artificial intelligence: A taxonomy and types. In Proceedings of the 57th Hawaii International Conference on system sciences. Hawaii, 2024. https://doi.org/10.24251/hicss.2024.546

Su, J., & Yang, W. (2023). Unlocking the power of ChatGPT: A framework for applying generative AI in education. *ECNU Review of Education*, *6*(3), 355–366. https://doi.org/10.1177/20965311231168423

Su, Y., Lin, Y., & Lai, C. (2023). Collaborating with ChatGPT in argumentative writing classrooms. *Assessing Writing*, *57*, 100752. https://doi.org/10.1016/j.asw.2023.100752

Sun, L., & Zhou, L. (2024). Does generative artificial intelligence improve the academic achievement of college students? A meta-analysis. *Journal of Educational Computing Research*, *62*(7), 1896–1933. https://doi.org/10.1177/07356331241277937

Susnjak, T., & McIntosh, T. R. (2024). ChatGPT: The end of online exam integrity? *Education Sciences (Basel)*, *14*(6), 656. https://doi.org/10.3390/educsci14060656

Takita, H., Kabata, D., Walston, S. L., Tatekawa, H., Saito, K., Tsujimoto, Y., & Ueda, D. (2024). Diagnostic performance comparison between generative AI and physicians: A systematic review and meta-analysis. medRxiv, 2024-01. https://doi.org/10.1101/2024.01.20.24301563

Tam, A. C. F. (2024). Interacting with ChatGPT for internal feedback and factors affecting feedback quality. *Assessment & Evaluation in Higher Education*, *50*(2), 219–235. https://doi.org/10.1080/02602938.2024.2374485

Tang, X., Tang, S., Ren, Z., & Wong, D. F. K. (2020). Psychosocial risk factors associated with depressive symptoms among adolescents in secondary schools in mainland China: A systematic review and meta-analysis. *Journal of Affective Disorders*, *263*, 155–165. https://doi.org/10.1016/j.jad.2019.11.118

Tantivejakul, N., Chantharasombat, J., & Kongpolphrom, W. (2024). Voices of the future: Exploring students' views on the use of GenAI in academic and professional PR writing. *LEARN Journal: Language Education and Acquisition Research Network*, *17*(2), 511–537. https://doi.org/10.70730/PVTD8302

Tardy, C. H. (1985). Social support measurement. *American Journal of Community Psychology*, *13*(2), 187–202. https://doi.org/10.1007/BF00905728

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*, 100270. https://doi.org/10.1016/j.caeai.2024.100270

Thomas, B. H., Ciliska, D., Dobbins, M., & Micucci, S. (2004). A process for systematically reviewing the literature: Providing the research evidence for public health nursing interventions. *Worldviews on Evidence-Based Nursing*, *1*(3), 176–184. https://doi.org/10.1111/j.1524-475X.2004.04006.x

* Uddin, S. J., Albert, A., Ovid, A., & Alsharef, A. (2023). Leveraging ChatGPT to aid construction hazard recognition and support safety education and training. *Sustainability (Basel)*, *15*(9), 7121. https://doi.org/10.3390/su15097121

Van den Noortgate, W., López-López, J. A., Marín-Martínez, F., & Sánchez-Meca, J. (2013). Three-level meta-analysis of dependent effect sizes. *Behavior Research Methods*, *45*(2), 576–594. https://doi.org/10.3758/s13428-012-0261-6

Vevea, J. L., & Woods, C. M. (2005). Publication bias in research synthesis: Sensitivity analysis using a priori weight functions. *Psychological Methods*, *10*(4), 428–443. https://doi.org/10.1037/1082-989X.10.4.428

Wang, F., Cheung, A. C., Neitzel, A. J., & Chai, C. S. (2024a). Does chatting with chatbots improve language learning performance? A meta-analysis of chatbot-assisted language learning. *Review of Educational Research*, 00346543241255621. https://doi.org/10.3102/00346543241255621

Wang, H., Dang, A., Wu, Z., & Mac, S. (2024b). Generative AI in higher education: Seeing ChatGPT through universities' policies, resources, and guidelines. *Computers and Education: Artificial Intelligence*, *7*, 100326. https://doi.org/10.1016/j.caeai.2024.100326

Wei, Q., Yao, Z., Cui, Y., Wei, B., Jin, Z., & Xu, X. (2024). Evaluation of ChatGPT-generated medical responses: A systematic review and meta-analysis. *Journal of Biomedical Informatics*, *151*, 104620. https://doi.org/10.1016/j.jbi.2024.104620

Wood, D., & Moss, S. H. (2024). Evaluating the impact of students' generative AI use in educational contexts. *Journal of Research in Innovative Teaching & Learning*, *17*(2), 152–167. https://doi.org/10.1108/JRIT-06-2024-0151

* Wu, C., Chen, L., Han, M., Li, Z., Yang, N., & Yu, C. (2024). Application of ChatGPT-based blended medical teaching in clinical education of hepatobiliary surgery. *Medical Teacher*. https://doi.org/10.1080/0142159X.2024.2339412

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

Xia, Q., Weng, X., Ouyang, F., Lin, T. J., & Chiu, T. K. (2024). A scoping review on how generative artificial intelligence transforms assessment in higher education. *International Journal of Educational Technology in Higher Education*, *21*(1), 40. https://doi.org/10.1186/s41239-024-00468-z

Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., Li, D., Jin, Y., & Gašević, D. (2024). Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology*, *55*(1), 90–112. https://doi.org/10.1111/bjet.13370

Yan, Z. (2022). *Student self-assessment as a process for learning*. Routledge.

Yan, Z., Lao, H., Panadero, E., Fernández-Castilla, B., Yang, L., & Yang, M. (2022). Effects of self-assessment and peer-assessment interventions on academic performance: A meta-

analysis. *Educational Research Review*, *37*, 100484. https://doi.org/10.1016/j.edurev.2022.100484

Yi, L., Liu, D., Jiang, T., & Xian, Y. (2024). The effectiveness of AI on K-12 students' mathematics learning: A systematic review and meta-analysis. *International Journal of Science and Mathematics Education*, 1–22. https://doi.org/10.1007/s10763-024-10499-7

Yu, H. (2023). Reflection on whether Chat GPT should be banned by academia from the perspective of education and teaching. *Frontiers in Psychology*, *14*, 1181712. https://doi.org/10.3389/fpsyg.2023.1181712

Yu, H. (2024). The application and challenges of ChatGPT in educational transformation: New demands for teachers' roles. *Heliyon*, *10*(2), e24289. https://doi.org/10.1016/j.heliyon.2024.e24289

Zhan, Y., & Yan, Z. (2025). Students' engagement with ChatGPT feedback: Implications for student feedback literacy in the context of generative artificial intelligence. *Assessment & Evaluation in Higher Education*, 1–14. https://doi.org/10.1080/02602938.2025.2471821

Zhan, Y., Yan, Z., Wan, Z. H., Wang, X., Zeng, Y., Yang, M., & Yang, L. (2023). Effects of online peer assessment on higher-order thinking: A meta-analysis. *British Journal of Educational Technology*, *54*(4), 817–835. https://doi.org/10.1111/bjet.13310

Zhang, X., Zhang, P., Shen, Y., Liu, M., Wang, Q., Gašević, D., & Fan, Y. (2024). A systematic literature review of empirical research on applying generative artificial intelligence in education. *Frontiers of Digital Education*, *1*(3), 223–245. https://doi.org/10.1007/s44366-024-0028-5

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

* Zhou, W., & Kim, Y. (2024). Innovative Music Education: An empirical assessment of ChatGPT-4's impact on student learning experiences. *Education and Information Technologies*, *29*(16), 20855–20881. https://doi.org/10.1007/s10639-024-12705-z

Zirar, A. (2023). Exploring the impact of language models, such as ChatGPT, on student learning and assessment. *The Review of Education*, *11*(3), e3433. https://doi.org/10.1002/rev3.3433

Author Biographies

Jiahe GU is a PhD student at the Department of Curriculum and Instruction at the Education University of Hong Kong. She obtained her MA degree from University College London (with distinction) and her BA degree from The Chinese University of Hong Kong, Shenzhen. Her expertise and interest lie in formative assessment, feedback, and Teaching English as a second/foreign language.

Professor Zi YAN is a RGC Senior Research Fellow and the Head of the Department of Curriculum and Instruction at the Education University of Hong Kong. His publications and research interests focus on two related areas, i.e., educational assessment in the school and higher education contexts with an emphasis on student self-assessment; and Rasch measurement, in particular its application in educational and psychological research. He is the author of Student self-assessment as a process for learning

(Routledge, 2022) and Applying the Rasch model: Fundamental measurement in the human sciences (Routledge, 2020).