International **R**eview of **R**esearch in **O**pen and **D**istributed **L**earning
Volume 27, Number 1

February – 2026

A Meta-Analysis of ChatGPT's Influence on Learning Achievement

Min Young Doo[1] and Yeonjeong Park[2*]
*[1]College of Education, Kangwon National University; [2]College of Education, Jeonbuk National University; *Corresponding Author*

Abstract

This meta-analysis synthesized empirical findings on the influence of ChatGPT on learning achievement. An electronic database search using Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines was conducted with relevant keywords to identify eligible research studies published between November 2022 and December 2024. A total of 22 eligible publications that met our inclusion criteria were reviewed. ChatGPT had a moderate positive effect on learning achievement (g = 0.573), indicating its potential to enhance learning outcomes. Subgroup analysis revealed that ChatGPT had a larger effect on middle and high school students (g = 0.928) than on undergraduates (g = 0.538), although the difference was not statistically significant. This finding highlights the importance for instructors and educational practitioners to consider the applications of ChatGPT in middle and high school settings. No significant statistical differences were found among the cognitive, affective, and metacognitive domains. Given that nearly half the studies focused on the cognitive domain, it is important to diversify the application of generative AI across a variety of subjects in different learning domains. The most frequently used instructional approaches with ChatGPT applications were lectures (22.1%) and self-regulated learning (16.3%). The effect of ChatGPT was largest on self-regulated learning (g = 1.115), followed by case-based learning (g = 0.836), while its effect was smallest on game-based learning (g = 0.092). This study was conducted within two years of ChatGPT's emergence, limiting our ability to analyze a large number of publications. Nevertheless, this study offers meaningful implications for future research on the application of ChatGPT for educational purposes.

Keywords: ChatGPT, learning achievement, generative AI, artificial intelligence in education, AIED

Introduction

Artificial intelligence (AI) has been defined by Popenici et al. (2017) as "computing systems that are able to engage in human-like processes such as learning, adapting, synthesizing, self-correction, and the use of data for complex processing tasks" (p. 2). ChatGPT by OpenAI, a representative commercially available generative AI, has rapidly gained attention since its emergence in November 2022. It is expected to bring about fundamental and unprecedented changes in society. As generative AI is evolving at a fast speed and is becoming more user friendly (e.g., from ChatGPT 3.0 in 2022 to GPT-4o in 2024), far more people have begun to use it in their daily lives. Reuters (2024) reported that OpenAI estimated the number of weekly ChatGPT users would exceed 200 million by the summer of 2024. The popularity of OpenAI has spread to education fields as well. Many organizations, such as UNESCO (Sabzalieva & Valentini, 2023) and the Council of Europe (2022) released special reports or issue papers about the application of AI in education and expected change in the field. Growing academic interest in artificial intelligence in education (AIED) has also prompted many top-tier journals to publish special issues on generative AI applications in education. In addition, new journals have recently launched that exclusively focus on AIED.

Díaz and Nussbaum (2024) defined AIED as "an educational technology capable of detecting patterns in existing or in vivo data and making automatic instructional decisions that are developed or implemented for pedagogical purposes to enhance the teaching and learning process" (p. 2). As Gagne's taxonomy of learning delineates (Gagne & Briggs, 1984), students traditionally have been expected to memorize definitions and concepts, apply rules and principles, understand cause and effect, and solve well-defined or ill-defined problems in the learning process. In contrast, generative AI, such as ChatGPT, requires students to proactively ask questions instead of passively answering given questions. Rospigliosi (2023) described the interaction or conversation between ChatGPT and its users, where a thread of user questions and ChatGPT responses enables an interactive learning process. This process encourages students to ask follow-up questions and challenge the responses of ChatGPT leading to enhanced learning engagement. This is the most distinctive feature of generative AI compared to other instructional media or intelligent tutoring systems (ITS) that have been adopted in education (Lo et al., 2024).

Educators have high expectations that generative AI can facilitate personalized learning, characterized by tailoring learning experiences for individuals and flexible learning paths (Gunawardena et al., 2024). Generative AI is also expected to accelerate the shift toward a learner-centered paradigm (Lee & Moore, 2024) and enhance learning effectiveness and efficiency (Alneyadi & Wardat, 2023; Hsu, 2024; Teng, 2024). UNESCO delineated potential areas for the application of AI in higher education, including teaching and learning, research, administration, and community engagement (Sabzalieva & Valentini, 2023). Crompton and Burke (2023) identified applications of AI in higher education with a more focused emphasis on teaching and learning, including assessment/evaluation, prediction, AI assistants, ITS, and managing student learning. Similarly, Díaz and Nussbaum's (2024) findings included ITS, prediction, diagnostics, adaptive systems, and assessment. These results imply that generative AI including ChatGPT is expected to have an impact across all areas of education.

Despite the high expectations of generative AI for pedagogical purposes, educators are also concerned about side effects and negative influences of ChatGPT applications on learning. UNESCO raised ethical concerns about using ChatGPT in education, such as academic integrity, cognitive bias, accessibility, and gender and diversity issues (Sabzalieva & Valentini, 2023). In their systematic review, Lo et al.

(2024) found that AI may reduce critical thinking and lead to overreliance on ChatGPT, contributing to student disengagement. They also noted an increase in plagiarism and cheating due to the use of ChatGPT in education. Dikilitaş et al. (2024) also raised concerns that excessive use of ChatGPT in writing or problem-solving activities may lower students' self-regulated learning. Many researchers have recently conducted rigorous research to examine the effects of ChatGPT on learning outcomes with high expectations of using ChatGPT for learning.

The purpose of this study was to synthesize current empirical research findings on the influence of ChatGPT on learning achievement using a meta-analysis method. Wu and Yu (2024) criticized the lack of meta-analyses synthesizing quantitative studies on the effects of generative AI, in contrast to the dozens of systematic review studies and substantial number of conceptual papers on the same theme (e.g., Crompton & Burke, 2023; Díaz & Nussbaum, 2024; Jeon et al., 2023).

A meta-analysis enables researchers to synthesize individual quantitative research findings and provide more robust and convincing conclusions (Borenstein et al., 2009). It may be premature to conduct a meta-analysis to synthesize empirical studies examining the influence of ChatGPT on learning achievement, as only two years had passed since the emergence of ChatGPT at the time this study was conducted. However, given that ChatGPT continues to evolve with growing attention and widespread use in education, it is worthwhile to conduct a meta-analysis to understand the status of ChatGPT application in education in the early stage of ChatGPT and to provide direction for future research.

Theoretical Framework

Driscoll (1993) defined learning as "a persisting change in human performance or performance potential" resulting from the learner's experience and interaction with the world. In line with this perspective, the present study conceptualizes learning achievement as measurable changes in students' attitudes, cognition, and behaviors. This broad definition acknowledges that achievement encompasses not only academic performance but also affective dispositions and metacognitive strategies that shape long-term learning potential.

Historically, the integration of new technologies into education has been accompanied by debates over their effects on learning outcomes. Saettler (1990) traced the evolution of instructional technology research alongside the emergence of new media, establishing the tradition of media comparison research. Within this paradigm, scholars have investigated whether novel technologies produce superior learning outcomes relative to earlier tools or traditional instruction. For instance, Steenbergen-Hu and Cooper (2014) synthesized findings on ITS and reported small to moderate effects on student achievement. Yet this tradition has also faced criticism. Warnick and Burbules (2007) cautioned that conflating media with instructional methods can lead to misleading conclusions, a critique that continues to resonate. More recently, Buchner and Kerres (2023) observed that media comparison research remains prevalent in fields such as augmented reality, suggesting that interest in technology–learning effects persists despite its contested status. As a cutting-edge technology, ChatGPT is increasingly integrated into teaching and learning, and there is growing interest in investigating its effects on learning achievement despite the controversy in media comparison research.

To better understand the potential influence of ChatGPT, we considered the following theoretical perspectives. Bloom's taxonomy and its revision by Anderson and Krathwohl (2001) highlighted the multiple dimensions of learning achievement, ranging from factual recall to higher-order reasoning and

self-regulation. Cognitive load theory (Sweller, 1988) has been supported by several recent studies (e.g., Becker et al., 2025; Martin et al., 2025; Patac & Patac, 2025), which have suggested that ChatGPT may reduce extraneous cognitive demands by providing immediate explanations, thereby allowing learners to focus on deeper processing. From a sociocultural perspective (Vygotsky, 1978), ChatGPT can be viewed as a form of digital scaffolding that supports learners in their zone of proximal development through interactive prompts and adaptive dialogue. Finally, frameworks such as the Technology Acceptance Model (Davis, 1989) and the Unified Theory of Acceptance and Use of Technology (Venkatesh et al., 2003) explain how learners' perceptions of usefulness and ease of use shape their willingness to integrate ChatGPT into their study practices, thereby influencing whether its affordances translate into measurable achievement gains. Taken together, these perspectives form a conceptual foundation for the present meta-analysis. While this study did not test a single overarching theory, the combination of media comparison traditions, educational taxonomies, cognitive and sociocultural learning theories, and technology adoption frameworks provides a multifaceted lens for interpreting the synthesized empirical findings on ChatGPT's influence on learning achievement.

Previous Reviews Examining ChatGPT's Influence on Learning

Several meta-analyses have recently examined the impact of ChatGPT and related AI chatbots on student learning, yet their findings reveal both promising effects and important limitations. For example, Deng et al. (2025) conducted one of the most comprehensive reviews, synthesizing 69 experimental studies published between 2022 and 2024. They reported strong positive effects of ChatGPT on academic achievement, affective states, and higher-order thinking, while finding weaker or inconsistent results for self-efficacy and mental effort. Their analysis further indicated that contextual factors such as intervention setting and duration moderated the outcomes, with classroom-based and longer interventions producing stronger effects. However, their study was restricted to experimental designs, thereby excluding correlational evidence and limiting the scope of generalization.

Building on this foundation, Wu and Yu (2024) expanded the focus to AI chatbots more broadly, analyzing 24 empirical studies across diverse educational settings. Their results also indicated large effect sizes for performance, motivation, self-efficacy, and interest, but a negative association with anxiety. Unlike Deng et al. (2025), Wu and Yu did not examine how instructional approaches or learning domains might shape these outcomes, which makes it difficult to translate their findings into concrete pedagogical practices. Similarly, Sun and Zhou (2024) narrowed their scope to college students, showing medium overall effects of generative AI on achievement and highlighting the effectiveness of text-based and independent learning conditions. Yet their review excluded younger learners and did not clarify how differences across school levels or disciplines might influence the impact of generative AI. Together, these three meta-analyses underscore the educational potential of ChatGPT and related tools, but each is constrained by its limited scope of participants, contexts, or moderating variables.

Beyond meta-analytic evidence, a number of systematic reviews have provided complementary insights into how ChatGPT and generative AI are being used in education. For instance, Lo et al. (2024) synthesized 72 studies on student engagement and found mixed outcomes, with relatively stronger evidence for behavioral engagement but weaker or inconsistent patterns for cognitive and affective engagement. Their review illustrated the breadth of ChatGPT applications but, lacking effect-size estimates, could not quantify the magnitude of such effects. Similarly, Dikilitaş et al. (2024) reviewed a small number of early empirical studies in higher education and identified themes such as skill development, feedback, motivation, and ethical concerns. Although valuable, their review was

necessarily limited by the short timeframe since ChatGPT's release. At the K–12 level, Díaz and Nussbaum (2024) offered a theoretically grounded review of 183 AI application studies using the Human-Centered AI framework, classifying interventions by learning theories such as constructivism and experiential learning. While their approach has enriched theoretical understanding, it does not provide quantitative evidence on learning outcomes. Taken together, these reviews highlight the rapid growth of ChatGPT-related research but also indicate the absence of systematic synthesis across learner groups, learning domains, and instructional approaches.

Earlier work on ITS has provided additional historical context. Steenbergen-Hu and Cooper's (2014) meta-analysis revealed that ITS had small to moderate effects on undergraduates' academic learning, with stronger outcomes compared to self-reliant learning but weaker results relative to human tutoring. Although situated in a pre-generative AI era, their findings point to the importance of considering both instructional context and comparator conditions when evaluating AI-based educational technologies. This insight remains relevant for assessing ChatGPT's role in contemporary classrooms.

Collectively, the existing body of reviews has provided important but partial insights. Meta-analyses have demonstrated substantial overall effects but have been often constrained by narrow participant groups, selective outcome measures, or limited consideration of moderating factors. Systematic reviews have deepened theoretical perspectives and mapped emerging trends but lack the empirical synthesis needed to quantify impact. These limitations underscore the need for a broader and more integrative meta-analysis that synthesizes recent empirical studies on ChatGPT, incorporates diverse learner populations, and explicitly examines the moderating effects of instructional approaches, learning outcomes, and research methods. The present study aims to address these gaps by offering a comprehensive meta-analysis of ChatGPT's influence on learning achievement across educational contexts.

Method

Given that the purpose of this study was to examine the effects of ChatGPT on learning achievement, we formulated the following research questions:

1. What are the overall effects of ChatGPT on learning achievement?
2. Are the effects of ChatGPT on learning achievement moderated by the school levels of participants?
3. How do the types of learning outcomes moderate the effects of ChatGPT on learning achievement?
4. Are the effects of ChatGPT on learning achievement moderated by instructional approaches?
5. Are the effects of ChatGPT on learning achievement moderated by research methods?

Literature Search

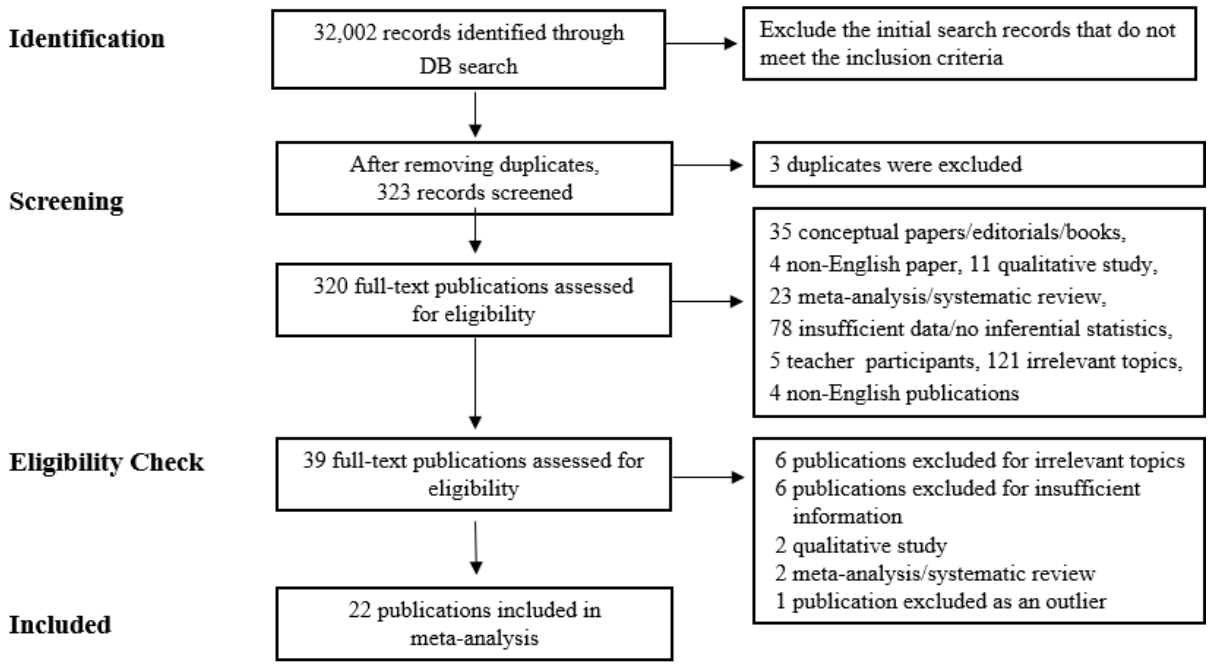
To search for eligible articles in an effective and efficient manner, we set five inclusion criteria satisfying the purpose of this research. The studies must have: (a) examined the effects of ChatGPT on learning achievement; (b) been published between November 2022 (the release of ChatGPT) and December 2024 (the time of the literature search); (c) adopted a rigorous research design; (d) reported quantitative learning outcomes (i.e., inferential statistics); and (e) been written and published in English. Exclusion criteria included qualitative studies (e.g., Garcia-Varela et al., 2025), conceptual

papers, meta-analysis or systematic reviews (i.e., secondary data analysis), publications that reported insufficient data for calculating effect sizes, and studies published in languages other than English.

We conducted an electronic database search using Web of Science, Google Scholar, Education Source (EBSCOhost), ERIC (ProQuest), PsycINFO, and JSTOR for Dissertations & Theses. We chose these databases because they cover most publications in education and many previous meta-analysis studies have used these electronic databases to search for relevant literature (Coban et al., 2022; Schoenherr et al., 2024). The keywords used for the search included "ChatGPT," "generative AI," "AIED," "artificial intelligence in education," "learning achievement," "learning outcomes," and their combinations. Using Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA), the initial database search yielded 32,002 publications, which were reduced to 320 publications after excluding redundant or ineligible studies that did not satisfy the inclusion criteria (see Figure 1). We conducted a full-text review of 320 publications and examined the eligibility of each publication. Through this process, 39 relevant samples remained. In addition, in the full-text review process, six samples were not included due to irrelevance to the topic, two samples were excluded for qualitative research, two samples were systematic reviews or meta-analyses, and another seven samples were excluded for insufficient information. Thus, a total of 86 independent studies (i.e., k = 86) in 22 publications were included for analysis.

Figure 1

Search and Exclusion Process



Note. DB = database. Adapted from *PRISMA Flow Diagram*, by PRISMA, 2020 (https://www.prisma-statement.org/prisma-2020-flow-diagram). CC BY 4.0.

Information Gathering and Extracting Information From Studies

From the 22 publications, 86 independent studies were coded as we extracted information related to four types of variables: independent variables, moderating variables, dependent variables, and other

variables (see Appendix). The extracted information was coded using a coding scheme in an Excel file. As illustrated in Table 1, we examined the effects of the moderating variables on the relationship between the independent and dependent variables, including the types of learning achievement, school levels of learners, learning disciplines, instructional approaches, and research methods.

Table 1

Coding Information for Meta-Analysis

Variable type	Categories	Coding information
Independent	Implementation of ChatGPT	ChatGPT versions
Moderating	School level	PreK–12, elementary, middle, and high school, higher education, adult learners
	Instructional approach	Direct instruction, PBL, flipped learning, team projects/collaborative learning, independent study
	Learning discipline	Languages, mathematics, STEM, humanities, business, social science, natural science, education, others, and unspecified
	Learning environment	Online learning, traditional classrooms, blended learning including flipped learning
Dependent	Learning outcomes	Affective, cognitive, metacognitive, behavioral (e.g., participation)
		Effect sizes: means and standard deviation, *t*-values, *F*-values, *r*, *p*-values, sample sizes
Other	Publication details	Title, author, year, journals, countries of publication
	Publication type	Journal articles, proceedings, unpublished dissertations or thesis, technical reports
	Research design	Experimental, quasi-experimental, one-group pretest-posttest design, and non-experimental (i.e., correlational studies)

Note. PBL = Problem-based learning; STEM = Science, technology, engineering, and mathematics

We coded the samples separately using the coding scheme detailed in Table 1. After discussing the coding scheme, we reached 91.7% intercoder agreement for the initial coding. Disagreements regarding coding were resolved through discussion of each case, after which we reached full consensus.

A brief description of the studies included in this meta-analysis is as follows. Among the 22 publications, five papers were published in 2023 (22.7%), and 17 papers (77.3%) were published in 2024. In terms of publication types, 21 publications were peer-reviewed journal articles, and one was a conference proceeding (i.e., Xue et al., 2024). These publications were published in 12 countries. Taiwan was the most productive country (seven studies, 31.8%), followed by China (four studies, 18.2%) and Turkey (two studies, 9.1%). Several countries published one study, including Arab Emirates, Czech, Ghana,

Hong Kong, Republic of Korea, Macau, Mexico, Norway, and the USA. Sample sizes in the publications ranged from 31 to 269, with an average sample size of 86.6.

Since different research designs were used in the publications, we examined the descriptive data for each effect size (i.e., k = 86), rather than by publication. We adopted Lo et al.'s (2024) category of research design, which is based on Creswell's (2012) classification. Among the research designs, the quasi-experimental design was the most frequently adopted (48 studies, 55.8%). The pretest-posttest experimental design with no control group was used in 18 studies (20.9%), while 10 studies (11.6%) employed a true-experimental design (i.e., pretest-posttest experimental design with random assignment). Additionally, two studies were correlational studies (2.3%), and eight studies (9.3%) used an experimental design without a pretest. Random assignment of participants was used in 38 studies (44.2%). There was a wide range of participants including elementary students (2 studies, 2.3%), middle and high school students (9 studies, 10.5%), undergraduates (61 studies, 70.9%), and adults including adult learners and clinical teachers (14 studies, 16.3%). Pre-K–12 participants (i.e., kindergarteners) were not included in this analysis. It is notable that more than half of the studies included undergraduates as participants in the analysis.

Learning outcomes as the dependent variable of this study were coded based on the four types of learning outcomes suggested by Bloom's taxonomy of learning objectives (Anderson, & Krathwohl, 2001) and Doo et al. (2023): (a) affective, (b) cognitive, (c) meta-cognitive, and (d) behavioral learning outcomes (i.e., participation). The affective domain included emotional engagement (Liang et al., 2024; Teng, 2024), learning motivation (Li, 2023; Ng et al., 2024; Teng, 2024; Yilmaz & Yilmaz, 2023), flow (Chen & Hou, 2024), attitudes toward learning, an intention for continuous learning, learning satisfaction, frustration, or anxiety (Chen & Hou, 2024; Liao et al., 2024). The cognitive domain of learning included mental activities, such as understanding and applying (Alneyadi & Wardat, 2023; Hsu, 2024; Huesca et al., 2024; Ng et al., 2024; Svendsen et al., 2024; Zhou & Kim, 2024), analyzing concepts and theories, generating knowledge (Guo et al., 2023; Xue et al., 2024) or problem solving (Chang et al., 2024; Lee et al., 2024; Urban et al., 2024). The behavioral domain included frequencies of logins, participation in learning activities, learning time, and completion of courses. The meta-cognitive domain of learning included higher-order thinking (Lee et al., 2024), planning, monitoring, critical thinking (Chang et al., 2024; Chang & G.-H. Hwang, 2024; Chang & G.-J. Hwang, 2024; Essel et al., 2024), collective efficacy (Urban et al., 2024), self-evaluation, and applications of self-regulatory strategies (Lee et al., 2024; Liao et al., 2024). In our research, nearly half the studies examined learning outcomes in the cognitive domain (44 studies, 51.2%) followed by the affective domain (25 studies, 29.0%) and the meta-cognitive domain (17 studies, 19.8%). No studies assessed learning outcomes in the psychomotor domain.

We also reviewed the instructional approaches used in the application of ChatGPT in each study. Of the 86 studies, 26 did not report the instructional approaches (30.2%). The most frequently used instructional approaches were lecture (19 studies, 22.1%) and self-regulated learning (nine studies, 10.5%) followed by problem-based learning (12 studies, 13.9%) and game-based learning (eight studies, 9.3%). Other instructional approaches included project-based learning (six studies, 7.0%), flipped learning (three studies, 3.5%), and case-based learning (three studies, 3.5%).

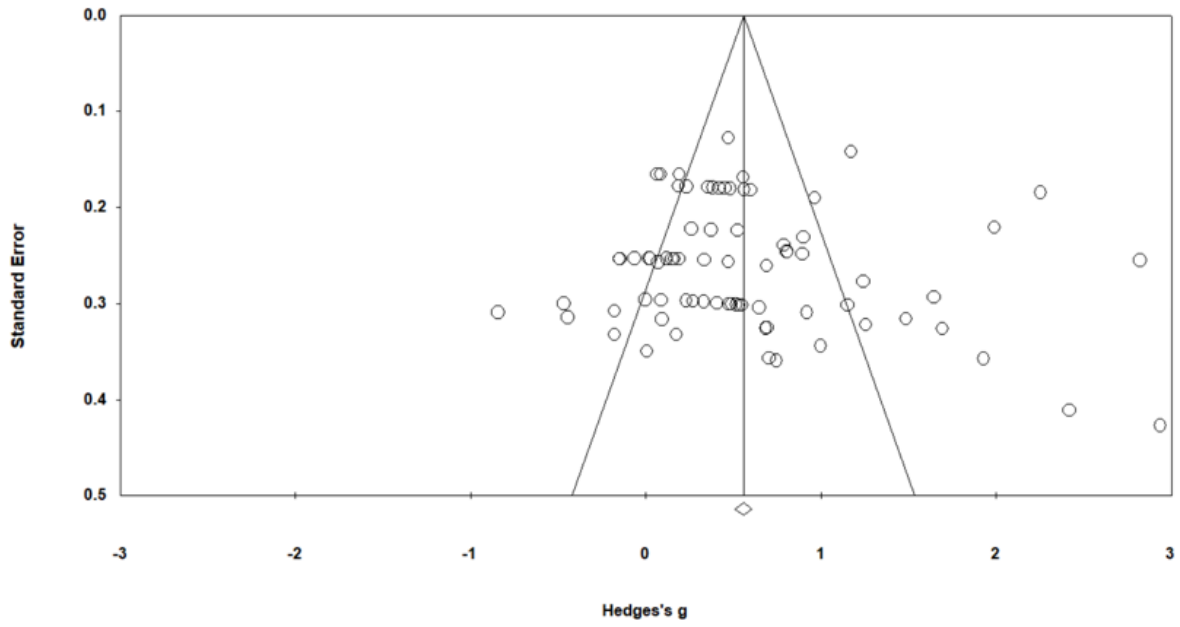
Analysis and Integration of Outcomes

We adopted Cohen's (1988) criteria to interpret the magnitude of effect sizes: 0.2 for a small effect size, 0.5 for a medium effect size, and 0.8 for a large effect size. We followed Higgins and Thompson's (2002) suggestions for interpreting I^2 scores: less than 30% would indicate mild heterogeneity, while more than 50% would indicate substantial heterogeneity. Depending on the heterogeneity of the effect sizes, we decided to use a random effects model. Among the different types of effect sizes, this study used Hedge's g as the type of effect size because more studies examined the group differences (e.g., pretest-posttest control group) than the relationship between the independent variables and dependent variables (e.g., correlation studies) in the analysis. In addition, because previous meta-analyses (e.g., Deng et al., 2025; Steenbergen-Hu & Cooper, 2014; Sun & Zhou, 2024) have used Hedge's g, it allowed us to easily compare our findings to their findings.

Publication bias was assessed using a funnel plot of the effect size and Egger's regression test. The funnel plot of effect size is slightly skewed to the right, as illustrated in Figure 2. Therefore, we double-checked for the absence of publication bias using Egger's regression tests. The results confirmed that publication bias was not detected (p = .59, t = .53; Egger et al., 1997). The meta-analysis was performed using Comprehensive Meta-Analysis (CMA Version 4.0) and IBM SPSS Statistics (Version 28.0).

Figure 2

Funnel Plot of Effect Size



Results

Overall Effects of ChatGPT on Learning Achievement

We synthesized the influences of ChatGPT on learning achievement using CMA software. The effect sizes ranged from g = -0.84 (Guo et al., 2023) to g = 2.94 (Teng, 2024), with the prediction interval [-.47, 1.61]. Statistics indicated substantial heterogeneity across the studies, Q = 511.1, df = 85, p < .001, I^2 = 83.37. Hence, we used a random-effects model to estimate the effect size and to examine the moderating effects of the variables (e.g., school level of learners, type of learning outcomes, type of

instructional approaches, and research method) using sub-group analysis. The overall effect size of the influence of ChatGPT on learning achievement is shown in Table 2. According to Cohen's (1988) criteria for effect size, the overall effect size of this study was medium-sized.

Table 2

Estimation of the Overall Effect Size Using the Random-Effects Model

Model	*k*	*N*	ES(*g*)	*SE*	Variance	95% CI	*Z* (2-tailed)	*p*
Random	6,992	86	0.573	0.063	0.004	[0.451, 0.696]	9.170	< .001

Note. ES = effect size.

Sub-Group Analysis Based on School Levels of Participants

We conducted a sub-group analysis based on the school levels of students (i.e., elementary school, middle and high school, undergraduate, adult learner, and clinical teacher) in the influence of ChatGPT on learning achievement. Results are illustrated in Table 3. The effect size of middle and high school students was larger than other learner groups. Elementary school students showed the smallest effect size. However, these results should be interpreted with caution due to the small number of effect sizes for elementary school studies (k = 2). The sub-group analysis results showed no significant differences among the effect sizes of sub-groups based on school levels, $Q_B(4) = 4.281$, $p = .369$. The effect size of middle and high school students was larger than it was with undergraduates. However, there were no statistical differences between the two groups. There were also no significant differences in the effect size between adult learners and clinical teachers in the same age range, despite their different roles and responsibilities in using ChatGPT.

Table 3

Effect of Learner Type on Influence of ChatGPT on Learning Achievement

Learner type	*k*	*N*	ES(*g*)	*SE*	95% CI	*Z* (2-tailed)	*p*
Adult learner	9	526	0.452	0.196	[0.067, 0.837]	2.301	.021
Clinical teacher	5	370	0.625	0.256	[0.123, 1.128]	2.439	.015
Elementary school	2	78	0.388	0.433	[-0.461, 1.237]	0.896	.370
Middle and high school	9	940	0.928	0.191	[0.554, 1.302]	4.861	.000
Undergraduate	61	5,088	0.538	0.075	[0.391, 0.684]	7.200	.000

Note. ES = effect size; CI = confidence interval.

Sub-Group Analysis Based on the Type of Learning Outcome

Next, we conducted a sub-group analysis based on the type of learning outcome (i.e., affective, cognitive, or meta-cognitive) in the influence of ChatGPT on learning achievement. Results are shown in Table 4. The effect size in the affective domain was smaller than the cognitive domain and the meta-cognitive domain. However, the results showed no significant differences among the effect sizes of the sub-

groups, $Q_B(2) = 0.983$, $p = .612$. The 95% confidence intervals indicated that no type of learning outcome was statistically higher or lower than the others.

Table 4

Influence of ChatGPT on the Type of Learning Outcome

Learning Outcome	*k*	*N*	ES(*g*)	*SE*	95% CI	*Z*	*p*
Affective	27	1,808	0.481	0.112	[0.262, 0.700]	4.303	.000
Cognitive	46	3,757	0.612	0.085	[0.446, 0.779]	7.205	.000
Meta-cognitive	13	1,427	0.619	0.153	[0.319, 0.919]	4.042	.000

Note. ES = effect size. CI = confidence interval.

Sub-Group Analysis Based on Type of Instructional Approach

We compared the effect sizes of the influence of ChatGPT on learning achievement with different instructional approaches (i.e., case-based, flipped learning, game-based, lecture, problem-based, project-based, and self-regulated learning) using sub-group analysis. Table 5 illustrates that self-regulated learning yielded the largest effect size, followed by case-based learning. The smallest effect size was observed with game-based learning, which was not statistically significant. The results indicate significant differences between instructional approaches, $Q_B(7) = 20.01$, $p < .05$.

Table 5

Influence of ChatGPT on the Type of Instructional Approach

Instructional approach	*k*	*N*	ES(*g*)	*SE*	95% CI	*Z*	*p*
Case-based	3	210	0.836	0.327	[0.196, 1.477]	2.560	.010
Flipped learning	3	416	0.626	0.315	[0.008, 1.244]	1.985	.047
Game-based	8	444	0.092	0.204	[-.0308, 0.491]	0.449	.653
Lecture	19	2,180	0.771	0.126	[0.524, 1.017]	6.127	.000
Not reported	26	1,699	0.343	0.114	[0.119, 0.567]	2.998	.003
Problem-based	12	1,325	0.589	0.160	[0.276, 0.902]	3.685	.000
Project-based	6	279	0.553	0.246	[0.071, 1.035]	2.247	.025
Self-regulated learning	9	439	1.115	0.201	[0.720, 1.510]	5.533	.000

Note. ES = effect size. CI = confidence interval.

Sub-Group Analysis Based on Research Method

Using sub-group analysis, we examined the moderating effect of research method (i.e., correlational studies, pretest-posttest control group design, mixed-design, quasi-experimental design, and posttest only design). As Table 6 illustrates, there were significant differences between sub-groups based on the research methods, $Q_B(4) = 18.474$, $p < .05$. An experimental design with posttest only was not statistically significant since its 95% confidence interval included zero. The effect size of correlational

studies was significantly larger than mixed method and quasi-experimental studies. However, there was no significant difference between correlational studies and true experimental studies (i.e., pretest-posttest control group studies).

Table 6

Influence of ChatGPT on Research Method

Research method	*k*	*N*	ES(*g*)	*SE*	95% CI	*Z*	*p*
Correlational	2	538	1.698	0.353	[1.007, 2.389]	4.814	.000
True experimental	18	1,769	0.762	0.126	[0.516, 1.009]	6.067	.000
Mixed methods	10	1,200	0.520	0.160	[0.207, 0.833]	3.257	.001
Post-test only	8	469	0.113	0.194	[-0.268, 0.494]	0.580	.580
Quasi-experimental	48	3,016	0.529	0.079	[0.374, .0683]	6.704	.000

Note. ES = effect size. CI = confidence interval.

Discussion

The purpose of this meta-analysis was to synthesize current empirical studies examining the influence of the ChatGPT application on learning achievement. We found a steep increase in the number of publications investigating the influence of ChatGPT on learning achievement, increasing from five papers in 2023 (22.7%) to 17 papers in 2024 (77.3%), which represents a nearly threefold increase in publications over 2 years. This finding supports Deng et al.'s (2025) observation of the increasing trend in ChatGPT-related publications. It also reveals that researchers' interest in the influence of ChatGPT on learning achievement in academia is growing rapidly, and this trend is expected to continue with the extensive adoption of ChatGPT in schools. In terms of the location of the publications, we found that among the 12 countries represented, Taiwan was the most productive (31.8%), followed by mainland China (18.2%). Deng et al. (2025) analyzed the productivity of publications by continent, not by nation, and found also that Asia (approximately 71.0%) was the most productive continent in their research. In addition, Lo et al.'s (2024) study highlighted that more than half the studies on AI applications for educational purposes were published in Asia (58.3%). These findings and those of two other studies suggest that researchers in Asia are more actively exploring the educational applications of ChatGPT compared to those in other continents.

Regarding the first research question, the overall effect size of the influence of ChatGPT on learning achievement was medium (g = 0.573). See Table 2. This supports the findings of Sun and Zhou (2024; g = 0.533) and Deng et al. (2025; g = 0.712). Our results were smaller than those of Wu and Yu (2024), who reported a large effect size (d = 0.964). However, ours were larger than Steenbergen-Hu and Cooper's (2014) effect size (g = 0.32), which estimated the influence of ITS on undergraduates' academic learning. Nevertheless, the medium effect size of ChatGPT's impact on learning achievement in this study shows promise for use of the application. This confirms the findings of previous meta-analytic and many empirical studies indicating that ChatGPT has the potential to improve learning achievement. These results can be interpreted through the lens of cognitive load theory (Sweller, 1988). By reducing extraneous demands through immediate explanations, ChatGPT may enable learners to

concentrate on germane processing, which helps explain the consistently positive effects observed in this and prior studies.

Related to the second research question, most participants in the studies we examined were undergraduates (70.9%), followed by middle and high school students (10.5%), and adult learners, including clinical teachers (16.3%). This distribution of was similar to that of Deng et al.'s (2025) research (undergraduates: 84.1%; K–12 students: 14.5%) and Lo et al.'s (2024) systematic review results (undergraduates: 79.2%; secondary school: 8.3%) These three studies indicate that the research on ChatGPT's influence on learning achievement has been more actively conducted in higher education than in K–12 settings.

This observation should be considered in light of the effect size of undergraduates versus K–12 students. Table 3 shows the results of the sub-group analysis based on participants' school level, indicating that the effect size of middle and high school students (g = 0.928) was larger than the effect size of undergraduates (g = 0.538) despite the non-statistical differences between the two groups. Despite the larger effect size for middle and high school students, few studies have examined the influence of ChatGPT on learning achievement in these settings. This discrepancy suggests that more research is needed in this area, and instructors and educational practitioners should work to further implement ChatGPT in middle and high school environments.

However, the research findings of this study are contradictory to that of Wu and Yu (2024), who reported a large effect size of undergraduates (d = 1.079) and a small effect size of secondary students (d = 0.214, *ns*). Deng et al. (2025) also reported that the effect size for undergraduates (g = 0.754) was larger than K–12 students (g = 0.547, p < .001), although there was no statistical difference between the two groups. Since they estimated the effect size for K–12 students, we estimated the effect size for K–12 in this study by combining elementary school students (k = 2) and middle and high school students (k = 9) for comparison purposes. The effect size for K–12 students (k =11) in this study was large, g = 0.84, 95% CI [0.498, 1.182], p < .001. Although it is smaller than the effect size for middle and high school students (g = 0.928), it was still larger than the effect size for undergraduates (g = 0.538). However, no statistical difference was found between K–12 students and undergraduates in the study. This finding supports Deng et al.'s (2025) finding of no moderating effects of school level on learning outcomes.

In terms of the type of learning outcome related to the third research question, we found that nearly half the studies focused on the cognitive domain (51.2%) followed by the affective domain (29.0%) and the meta-cognitive domain (19.8%). These results indicate that ChatGPT has been used extensively to help learners in activities such as memorizing terminology or knowledge (Hsu, 2024; Ng et al., 2024), problem-solving (Chang et al., 2024; Lee et al., 2024; Urban et al., 2024), computer programming (Xue et al., 2024), and argumentation skills (Guo et al., 2023). However, few studies examined learning outcomes in the meta-cognitive domain, such as higher-order thinking skills (Lee et al., 2024) and critical thinking or reflection skills (Chang et al., 2024; Chang & G.-H. Hwang, 2024; Chang & G.-J. Hwang, 2024; Essel et al., 2024). Thus, more ChatGPT applications should be developed that focus on learners' meta-cognitive skills as well as in the affective domain.

As shown in Table 4, the lack of a statistical difference among the three learning domains—cognitive (g = 0.612), affective (g = 0.481) and meta-cognitive (g = 0.619)—implies that students can benefit from using ChatGPT to achieve learning outcomes regardless of the learning domain. The large number of

ChatGPT applications focusing on the cognitive domain may be due to the heavy emphasis on the cognitive domain in the curriculum (e.g., knowledge, comprehension, application, analysis, synthesis, and evaluation; Bloom, 1956). Smith and Ragan (2005) pointed out that attitudinal learning (affective domain) and psychomotor skills (psychomotor learning domain) also include substantial cognitive components, such as the knowledge structure of learning content or know-how. Notably, no publication in this meta-analysis addressed the psychomotor skill domain. Educational researchers and practitioners should pay more attention to ChatGPT applications or other generative AI for a variety of subjects in different learning domains beyond the cognitive. This balanced influence across domains resonates with Anderson and Krathwohl's (2001) taxonomy, which emphasized that achievement encompasses not only cognitive skills but also affective dispositions and metacognitive strategies.

To answer the fourth research question, we also compared the effect sizes of ChatGPT's influence on learning achievement across different instructional approaches (case-based, flipped learning, game-based, lecture, problem-based, project-based, and self-regulated learning) using subgroup analysis. The most frequently used instructional approaches with ChatGPT applications were lecture (22.1%) and self-regulated learning (16.3%). The finding that lecture was the most frequently used instructional approach indirectly supports Deng et al.'s (2025) observation that most ChatGPT intervention settings were classrooms (86.96%). The seven instructional approaches coded for analysis were grouped into instructor-led approaches (e.g., lecture) and learner-centered approaches (e.g., case-based, flipped learning, game-based, problem-based, project-based, and self-regulated learning). Excluding studies that did not indicate instructional approaches (N = 26), we found that ChatGPT was used substantially more in learner-centered approaches (68.3%) than in traditional instructor-led approaches (31.7%). Given that ChatGPT is being adopted across various instructional approaches, it is important to investigate how ChatGPT is used in each approach and to develop customized instructional strategies to maximize learning outcomes.

In terms of the moderating effect of instructional approach, there were statistical differences among the various instructional approaches in terms of the influence of ChatGPT on learning achievement. These are shown in Table 5. The effect size for self-regulated learning was the largest effect size (g = 1.115) followed by case-based learning (g = 0.836), while the smallest effect size was game-based learning (g = .092, *ns*). This research finding supports Sun and Zhou's (2024) and Steenbergen-Hu and Cooper's (2014) results. Sun and Zhou (2024) reported that generative AI's influence on college students' academic achievement is most effective in the independent learning setting (g = 0.600) compared to cooperative learning (g = 0.328). Steenbergen-Hu and Cooper (2014) reported that ITS have a large effect size for self-regulated learning (self-reliant learning or no instruction; g = 0.86), but have a small effect size for classroom instruction (instructional learning; g = 0.37) where lectures are frequently used. The plausible reasons for the statistical differences in ChatGPT applications across different instructional approaches include whether the instructional methods give learners autonomy or the amount of structure for the learning activities (e.g., well-structured problems vs. poorly-structured problems). As Rospigliosi (2023) pointed out, learners benefit from ChatGPT when they ask questions using the application or interact with it as they progress in their learning. However, if students are expected to follow predefined learning paths, they may not feel the need to use ChatGPT, and we cannot expect it to improve learning outcomes.

This is why the effect size for self-regulated learning and problem-based learning was large, while the effect size for game-based learning was small. This finding implies that instructors need to understand

how students use ChatGPT in the learning process with different instructional methods and how learning activities should be designed to leverage the strengths of generative AI. These differences also resonate with self-regulated learning theory and technology acceptance perspectives (Davis, 1989; Venkatesh et al., 2003), which stress that learners' autonomy and their perceptions of usefulness critically mediate whether a technology such as ChatGPT could enhance achievement.

Regarding the last research question, most publications in this analysis adopted an experimental research design, including quasi-experimental and pretest-posttest control group studies, except for one publication (i.e., Dalgıç et al., 2024). It is promising that the research findings were derived from experimental research because experimental design studies provide stronger internal validity than correlational studies (Campbell & Stanley, 1963). We confirmed the moderating effect of research method (i.e., correlational studies, pretest-posttest control group design, mixed-design, quasi-experimental design, and posttest only design) on the influence of ChatGPT on learning achievement. See Table 6. The effect size for correlational studies (g = 1.698, p < .001) was larger than experimental design, such as pretest-posttest control group studies (g = 0.762, p < .001), and quasi-experimental studies (g = 0.529, p < .001). However, despite the large effect size for correlational studies, it should be noted that these studies, which are typically measured using self-reports such as surveys or questionnaires, carry the risk of common-method variance (Doo et al., 2023; Garger et al., 2019).

The limitations of this study have the potential to help readers better understand and interpret the results of this meta-analysis. First, this study analyzed only 22 publications, as ChatGPT was launched only 2 years before conducting this study. Given the growing academic interest in ChatGPT for educational purposes, future research will likely include a larger number of empirical studies. Second, since a meta-analysis is a secondary examination of individual experimental studies, our analysis was restricted to the data and information provided by the authors of each publication. This is an inherent limitation of a meta-analysis. Third, this study included only empirical research published in English. Our findings indicated that Asia was more productive in this area, and thus numerous studies have likely been published in local journals in Asian languages such as Chinese, Korean, or other regional languages. Future research extending this study may want to include publications in languages other than English. Finally, since the number of studies that met our inclusion criteria was limited, we did not exclude any based on quality. Future researchers are encouraged to conduct quality assessments to enhance validity.

Generative AI was not originally developed for teaching and learning. Over the past 2 years, innovators and early adopters in education have made bold attempts to integrate ChatGPT into teaching and learning with scholarly curiosity. This mirrors a recurrent pattern in the emergence of new instructional media throughout the history of educational technology (Reiser, 2001). The results of this study are promising, as the effect size for ChatGPT's influence on learning achievement was medium or moderate. The research findings give educational practitioners and researchers a solid rationale for implementing ChatGPT for educational purposes and exploring how students can learn with and from generative AI to achieve desirable learning outcomes across different instructional methods and learning environments. However, the application of ChatGPT itself does not guarantee successful learning outcomes. To fully leverage ChatGPT, educators need clear guidelines on how to integrate it as a scaffolding tool rather than a mere content provider, aligning with instructional objectives and learners' needs. Policymakers should also consider differentiated strategies across contexts, given the varying effects across educational levels and instructional approaches. To develop practical and effective

guidelines for ChatGPT applications in education, more empirical studies are needed to investigate the influence of ChatGPT on learning achievement in a variety of educational settings.

Declarations

- Funding: The authors declare that they have no funding source for this research.
- Ethical approval and informed consent statements: Not applicable.
- Conflict of interest: There were no conflicts of interest in this research.
- Data availability: The data used and/or analyzed in the current study are available from the corresponding author upon reasonable request.

References

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

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives: Complete edition*. Addison Wesley Longman, Inc.

Becker, E., Wünsche, J., Veith, J. M., Schrader, J., Bitzenbauer, P. (2025). *From cognitive relief to affective engagement: An empirical comparison of AI chatbots and instructional scaffolding in physics education*. arXiv. https://doi.org/10.48550/arXiv.2508.06254

Bloom, B. S. (Ed.). (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain*. New York, NY: David McKay.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. John Wiley & Sons.

Buchner, J., & Kerres, M. (2023). Media comparison studies dominate comparative research on augmented reality in education. *Computers & Education*, 195, Article 104711. https://doi.org/10.1016/j.compedu.2022.104711

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Rand-McNally.

Chang, C.-C., & Hwang, G.-H. (2024). Promoting professional trainers' teaching and feedback competences with ChatGPT: A question-exploration-evaluation training mode. *Educational Technology & Society*, *27*(2), 405–421. https://doi.org/10.30191/ETS.202404_27(2).TP06

Chang, C.-C., & Hwang, G.-J. (2024). ChatGPT facilitated professional development: Evidence from professional trainers' learning achievements, self-worth, and self-confidence. *Interactive Learning Environments*, *33*(1), 883–900. https://doi.org/10.1080/10494820.2024.2362798

Chang, C.-Y., Yang, C.-L., Jen, H.-J., Ogata, H., & Hwang, G.-H. (2024). Facilitating nursing and health education by incorporating ChatGPT into learning designs. *Educational Technology & Society*, *27*(1), 215–230. https://doi.org/10.30191/ETS.202401_27(1).TP02

Chen, Y.-C., & Hou, H.-T. (2024). A mobile contextualized educational game framework with ChatGPT interactive scaffolding for employee ethics training. *Journal of Educational Computing Research*, *62*(7), 1737–1762. https://doi.org/10.1177/07356331241268505

Coban, M., Bolat, Y. I., & Goksu, I. (2022). The potential immersive virtual reality to enhance learning: A meta-analysis. *Educational Research Review, 36*, Article 100452. https://doi.org/10.1016/j.edurev.2022.100452

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

Council of Europe. (2022). *Artificial intelligence and education*. https://www.coe.int/en/web/education/artificial-intelligence

Creswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). Pearson.

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

Dalgıç, A., Yaşar, E., & Demir, M. (2024). ChatGPT and learning outcomes in tourism education: The role of digital literacy and individualized learning. *Journal of Hospitality, Leisure, Sport & Tourism Education, 34,* Article 100481. https://doi.org/10.1016/j.jhlste.2024.100481

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340. https://doi.org/10.2307/249008

Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2025). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. *Computers & Education, 227,* Article 105224. https://doi.org/10.1016/j.compedu.2024.105224.

Díaz, B., & Nussbaum, M. (2024). Artificial intelligence for teaching and learning in schools: The need for pedagogical intelligence. *Computers & Education, 217,* Article 105071. https://doi.org/10.1016/j.compedu.2024.105071

Dikilitaş, K., Klippen, M. I. F., & Keles, S. (2024). A systematic rapid review of empirical research on students' use of ChatGPT in higher education. *Nordic Journal of Systematic Reviews in Education, 2,* 103–125.

Doo, M. Y., Zhu, M., & Bonk, C. J. (2023). MOOC learners' self-directed learning and learning outcomes: A meta-analysis. *Distance Education, 44*(1), 86-105. https://doi.org/10.1080/01587919.2022.2155618

Driscoll, M. P. (1993). *Psychology of learning for instruction* (2nd ed.). Allyn & Bacon.

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

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

Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., & Gašević, D. (2024). 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

Gagne, R. M. & Briggs, L. J. (1984). *Principles of instructional design*. Holt, Rinehart, and Winston.

Garcia-Varela, F., Beckerman, Z., Nussbuam, M., & Mendoza, M. (2025). Reducing interpretative ambiguity in an educational environment with ChatGPT. *Computers & Education, 225,* Article 105182. https://doi.org/10.1016/j.compedu.2024.105182

Garger, J., Jacques, P. H., Gastle, B. W., & Connolly, C. M. (2019). Threats of common method variance in student assessment of instruction instruments. *Higher Education Evaluation and Development, 13*(1), 2–17. https://doi.org/10.1108/HEED-05-2018-0012

Gunawardena, M., Bishop, P., & Aviruppola, K. (2024). Personalized learning: The simple, the complicated, the complex and the chaotic. *Teaching and Teacher Education, 139*, Article 104429. https://doi.org/10.1016/j.tate.2023.104429

Guo, K., Zhong, Y., Li, D., & Chu, S. K. W. (2023). Effects of chatbot-assisted in-class debates on students' argumentation skills and task motivation. *Computers & Education*, *203,* Article 104862. https://doi.org/10.1016/j.compedu.2023.104862

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

Hsu, M.-H. (2024). Mastering medical terminology with ChatGPT and Termbot. *Health Education Journal*, *83*(4), 352–358. https://doi.org/10.1177/0017896923119737

Huesca, G., Martínez-Trevino, Y., Molina-Espinosa, J. M., Sanromán-Calleros, A. R., Martínez-Román, R., Cendejas-Castro, E., & Bustos, R. (2024). Effectiveness of using ChatGPT as a tool to strengthen benefits of the flipped learning strategy. *Educational Science, 14,* Article 660. https://doi.org/10.3390/educsci14060660

Jeon, J., Lee, S. & Choe, H. (2023). Beyond ChatGPT: A conceptual framework and systematic review of speech-recognition chatbots for language learning. *Computers and Education, 206,* Article 104898. https://doi.org/10.1016/j.compedu.2023.104898

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,* Article 16. https://doi.org/10.1186/s41239-024-00447-4

Lee, S. S., & Moore, R. L. (2024). Harnessing generative AI (GenAI) for automated feedback in higher education: A systematic review. *Online Learning, 28*(3), 82–106. https://doi.org/10.24059/olj.v28i3.4593

Li, H.-F. (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

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

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*, Article 105100. https://doi.org/10.1016/j.compedu.2024.105100

Martin, A. J., Collie, R. J., Kennett, R., Liu, D., Ginns, P., Sudimantara, L. B., ... & Rüschenpöhler, L. G. (2025). Integrating generative AI and load reduction instruction to individualize and optimize students' learning. *Learning and Individual Differences*, 121, 102723. https://doi.org/10.1016/j.lindif.2025.102723

Ng, D. T. K., Tan, C. W., & Leung, J. K. L. (2024). Empowering student self-regulated learning and science education through ChatGPT: A pioneering pilot study. *British Journal of Educational Technology, 55*(4), 1328–1353. https://doi.org/10.1111/bjet.13454

Nwana, H. S. (1990). Intelligent tutoring systems: An overview. *Artificial Intelligence Review, 4,* 251–277. https://doi.org/10.1007/BF00168958

Patac, L. P., & Patac Jr, A. V. (2025). Using ChatGPT for academic support: Managing cognitive load and enhancing learning efficiency–A phenomenological approach. *Social Sciences & Humanities Open*, 11, 101301. https://doi.org/10.1016/j.ssaho.2025.101301

Popenici, S. A. D., & Kerr, S. (2017). Exploring the impact of artificial intelligence on teaching and learning in higher education. *Research and Practice in Technology Enhanced Learning, 12*, Article 22. https://doi.org/10.1186/s41039-017-0062-8

PRISMA. (2020). *PRISMA flow diagram*. https://www.prisma-statement.org/prisma-2020-flow-diagram

Reiser, R. A. (2001). A history of instructional design and technology: Part I: A history of instructional media. *Educational Technology Research and Development, 49*(1), 53–64. https://doi.org/10.1007/BF02504506

Reuters. (2024, August 30). *OpenAI says ChatGPT's weekly users have grown to 200 million*. https://www.reuters.com/technology/artificial-intelligence/openai-says-chatgpts-weekly-users-have-grown-200-million-2024-08-29/

Rospigliosi, P. (2023). Artificial intelligence in teaching and learning: What questions should we ask of ChatGPT? *Interactive Learning Environments*, *31*(1), 1–3. https://doi.org/10.1080/10494820.2023.2180191

Sabzalieva, E., & Valentini, A. (2023). *ChatGPT and artificial intelligence in higher education: Quick start guide*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000385146

Saettler, P. (1990). *The evolution of American educational technology*. Libraries Unlimited.

Schoenherr, J., Strohmaier, A. R, & Schukajlow, S. (2024). Learning with visualizations helps: A meta-analysis of visualization intervention in mathematics education. *Educational Research Review, 45*, Article 100639. https://doi.org/10.1016/j.edurev.2024.100639

Smith, P. L., & Ragan, T. J. (2005). *Instructional design* (3rd ed.). Wiley.

Steenbergen-Hu, S., & Cooper, H. (2014). A meta-analysis of the effectiveness of intelligent tutoring systems on college students' academic learning. *Journal of Educational Psychology, 106*(2), 331–347. https://doi.org/10.1037/a0034752

Sun, L., & Zhou, L. (2024). Does generative artificial intelligence improve the academic achievement of colleges? A meta-analysis. *Journal of Educational Computing Research*, *62*(7), 1673-1713. https://doi.org/10.1177/07356331241277937

Svendsen, K., Askar, M., Umer, D., & Halvorsen, K. H. (2024). Short-term learning effect of ChatGPT on pharmacy students' learning. *Exploratory Research in Clinical and Social Pharmacy*, *15,* Article 100478. https://doi.org/10.1016/j.rcsop.2024.100478

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

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

Urban, M., Děchtěrenko, F., Lukavský, J., Hrabalová, V., Svacha, F., Brom, C., & Urban, K. (2024). ChatGPT improves creative problem-solving performance in university students: An experimental study. *Computers & Education, 215,* Article 105031. https://doi.org/10.1016/j.compedu.2024.105031

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly, 27*(3), 425–478. https://doi.org/10.2307/30036540

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Warnick, B. R., & Burbules, N. C. (2007). Media comparison studies: Problems and possibilities. *Teachers College Record*, *109*(11), 2483–2510. https://eric.ed.gov/?id=EJ820500

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

Xue, Y., Chen, H., Bai, G. R., Tairas, R., & Huang, Y. (2024). Does ChatGPT help with introductory programming? An experiment of using students ChatGPT in CS1. In A. Paiva, R. Abreu, K. Gama, & J. Siegmund (Chairs), *ICSE-SEET '24: Proceedings of the 46th international conference on software engineering: Software engineering education and training* (pp. 331–341). ACM. https://doi.org/10.1145/3639474.3640076

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

Zhang, K., & Aslan, A. B. (2021). AI technologies for education: Recent research and future directions. *Computers and Education: Artificial Intelligence, 2,* Article 100025. https://doi.org/10.1016/j.caeai.2021.100025

Zhou, W., & Kim, Y. (2024). Innovative music education: An empirical assessment of ChatGPT-4's impact on student learning experiences. *Education and Information Technology, 29,* 20855–2088. https://doi.org/10.1007/s10639-024-12705-z

Appendix

Overview of Selected Papers

Author(s) & Year	Participants	Instructional approach	Learning discipline	Domain of learning outcomes		Research design
Alneyadi & Wardat (2023)	K-12 students (Middle and high)	Lecture	Science	C	Cognitive achievement (knowing, applying, reasoning)	Mixed-methods
Chang & H. Hwang (2024)	Clinical teacher	Problem based learning	Health	C M	Learning achievement Critical thinking	Quasi-experimental
Chang & J. Hwang (2024)	Clinical teachers	Case-based learning	Education	C M A	Test scores, Critical thinking Self-worth, self-confidence	Quasi-experimental
Chang et al. (2024)	Adult learners	Problem based learning	Health	C M A	Problem solving Critical thinking Learning enjoyment	Quasi-experimental
Chen & Hou (2024)	Adult learners	Game-based learning	Ethics	C A	Learning achievement Flow antecedent, motivation, anxiety, flow experience, flow	Quasi-experimental
Dalgıç et al. (2024)	Undergraduates	Not reported	Tourism	C	Digital literacy, individualized learning	Correlational
Essel et al. (2024)	Undergraduates	Lecture	Research	M	Critical thinking skills, creative thinking, reflection, thinking skills	Mixed-methods
Fan et al. (2024)	Undergraduates	Not reported	Language	A	Interests/enjoyment, perceived competence, efforts, pressure/tension	Post only
Guo et al. (2023)	Undergraduates	Not reported	Language	C A	Argumentation skills Task motivation	Quasi-experimental
Huesca et al. (2024)	Undergraduates	Flipped learning	Computer science	C	Learning gain in computer science	Experimental (Pre-Post test)
Hsu (2024)	Undergraduates	Lecture	Medical	C	Learning achievement	Quasi-experimental

Author(s) & Year	Participants	Instructional approach	Learning discipline	Domain of learning outcomes		Research design
Lee et al. (2024)	Undergraduates	Self-regulated learning	Science	C	Knowledge construction Problem-solving	Experimental (Pre-Post test)
				M	Critical thinking, creativity	
Li (2023)	Undergraduates	Flipped learning	Education	C	Project performance	Quasi-experimental
				A	Learning motivation	
Liang et al. (2024)	K-12 students (Elementary)	Game-based learning	Interdisciplinary learning	C	Learning achievement	Quasi-experimental
				A	Emotional engagement	
Liao et al. (2024)	K-12 students (Middle and high)	Lecture	Biology	C	Test-scores	Quasi-experimental
				M	Cognitive strategies	
				A	Test anxiety, self-efficacy	
Ng et al. (2024	K-12 students (Middle and high)	Self-regulated learning	Science	C	Science knowledge	Quasi-experimental
				A	Motivation	
Svendsen et al. (2024)	Undergraduates	Self-regulated learning	Pharmacy	C	Knowledge test	Experimental (Pre-Post test)
Teng (2024)	Undergraduates	Self-regulated learning	Language	A	Learning motivation, learning engagement	Experimental (Pre-Post test)
Urban et al. (2024)	Undergraduates	Problem based learning	Creative thinking	C	Problem-solving	Experimental (Pre-Post test)
				M	Self-evaluation, mental effort	
				A	Self-efficacy, making the task interesting	
Xue et al. (2024)	Undergraduates	Project based learning	Computer science	C	UML diagram, Java Programing, Post-evaluation scores	Post test only
Yilmaz & Yilmaz (2023)	Undergraduates	Project based learning	Computer science	C	Computational thinking	Experimental (Pre-Post test)
				A	Self-efficacy, motivation	
Zhou & Kim (2024)	Undergraduates	Lecture	Music	C	Music knowledge	Quasi-experimental

Note. C = cognitive; A = affective; M = meta-cognitive; UML = unified modeling language


Athabasca
University


BY