Humanities & Social Sciences Communications

REVIEW

https://doi.org/10.1057/s41599-025-04787-y OPEN

The effect of ChatGPT on students' learning performance, learning perception, and higher-order thinking: insights from a meta-analysis

Jin Wang[1] & Wenxiang Fan[1,2]✉

As a new type of artificial intelligence, ChatGPT is becoming widely used in learning. However, academic consensus regarding its efficacy remains elusive. This study aimed to assess the effectiveness of ChatGPT in improving students' learning performance, learning perception, and higher-order thinking through a meta-analysis of 51 research studies published between November 2022 and February 2025. The results indicate that ChatGPT has a large positive impact on improving learning performance ($g = 0.867$) and a moderately positive impact on enhancing learning perception ($g = 0.456$) and fostering higher-order thinking ($g = 0.457$). The impact of ChatGPT on learning performance was moderated by type of course ($Q_B = 64.249$, $P < 0.001$), learning model ($Q_B = 76.220$, $P < 0.001$), and duration ($Q_B = 55.998$, $P < 0.001$); its effect on learning perception was moderated by duration ($Q_B = 19.839$, $P < 0.001$); and its influence on the development of higher-order thinking was moderated by type of course ($Q_B = 7.811$, $P < 0.05$) and the role played by ChatGPT ($Q_B = 4.872$, $P < 0.05$). This study suggests that: (1) appropriate learning scaffolds or educational frameworks (e.g., Bloom's taxonomy) should be provided when using ChatGPT to develop students' higher-order thinking; (2) the broad use of ChatGPT at various grade levels and in different types of courses should be encouraged to support diverse learning needs; (3) ChatGPT should be actively integrated into different learning modes to enhance student learning, especially in problem-based learning; (4) continuous use of ChatGPT should be ensured to support student learning, with a recommended duration of 4–8 weeks for more stable effects; (5) ChatGPT should be flexibly integrated into teaching as an intelligent tutor, learning partner, and educational tool. Finally, due to the limited sample size for learning perception and higher-order thinking, and the moderately positive effect, future studies with expanded scope should further explore how to use ChatGPT more effectively to cultivate students' learning perception and higher-order thinking.

[1] Jing Hengyi School of Education, Hangzhou Normal University, Hangzhou, China. [2] Chinese Education Modernization Research Institute of Hangzhou Normal University (Zhejiang Provincial Key Think Tank), Hangzhou, China. ✉email: wenxiang_fan1990@163.com

Introduction

Constructivist learning theory holds that students need to interact with the environment and mentally construct an understanding of knowledge in order for effective learning to occur (Piaget, 1970; Vygotskij and Hanfmann, 1966). With the development of modern information technologies, some researchers have attempted to use that technology to create intelligent learning environments that support students' construction of knowledge and facilitate the occurrence of effective learning (Rapti et al. 2023; Wang et al. 2024b; Zhang et al. 2022). ChatGPT is a newly popular technology that has garnered significant attention. Its powerful knowledge base, rapid feedback, and content generation can engage students and foster interaction with it, suggesting a very high potential for supporting student learning. To explore the effectiveness of ChatGPT in supporting student learning, some researchers have tried to introduce ChatGPT into students' learning practice.

Many studies have claimed that ChatGPT can improve students' learning performance, learning perception, and higher-order thinking (Jalil et al. 2023; Kasneci et al. 2023; Lu et al. 2024), but other studies have found that it does not significantly improve students' learning performance or learning perception (Escalante et al. 2023; Donald et al. 2024) and can even hinder students' learning performance, learning perception, and higher-order thinking (Niloy et al. 2023; Yang et al. 2025). Thus, there is as yet no unified conclusion on whether ChatGPT can significantly improve students' learning performance, learning perception, and higher-order thinking. Therefore, it is necessary to provide objective and scientific evidence for the overall impact of ChatGPT on student learning.

Meta-analysis can be used to comprehensively analyze the results of multiple independent research studies that are focused on the same research purpose, reveal the differences and commonalities among them, and thereby draw universal and regular conclusions (Lipsey and Wilson, 2001). Therefore, this study used the meta-analysis method to analyze the results of 51 (quasi) experimental studies published between November 2022 and February 2025 to clarify the impact of ChatGPT on students' learning performance, learning perception, and higher-order thinking, and the circumstances under which ChatGPT can be used well. Specifically, this study focuses on the following two research questions:

RQ1: How effective is ChatGPT in promoting student learning performance, learning perception, and higher-order thinking?

RQ2: How do study characteristics such as grade level, type of course, learning model, duration, the role played by ChatGPT in teaching and learning, and area of ChatGPT application moderate the impact of ChatGPT?

Literature review

ChatGPT in education. ChatGPT is a prime example of a generative artificial intelligence (GenAI) technology chatbot, launched by OpenAI in November 2022 (OpenAI, 2022). Several scholars have pointed out that ChatGPT possesses diverse functions that can enable it to provide support to students in a wide range of learning contexts (Hsu, 2024; Maurya, 2024; Nugroho et al. 2023).

Studies to date have primarily explored the educational applications of ChatGPT in higher, secondary, and primary education. In higher education, research has focused on leveraging ChatGPT to assist students in academic writing (Nugroho et al. 2024; Song and Song, 2023), facilitate the comprehension of complex knowledge concepts (Haindl and Weinberger, 2024; Sun et al. 2024; Zhou and Kim, 2024), and solve complex problems (Küchemann et al. 2023; Urban et al. 2024). At the level of secondary education, scholars have primarily explored the use of ChatGPT to assist students in solving difficult problems (Bitzenbauer, 2023; Levine et al. 2025; Yang et al. 2025) and support knowledge review (Wu et al. 2023). In primary education, studies have explored the use of ChatGPT in addressing students' learning challenges (Kotsis, 2024a, 2024b), assisting in optimizing learning plans (Almohesh, 2024), and generating personalized learning resources (Jauhiainen and Guerra, 2024).

In terms of type of courses, scholars have explored the educational applications of ChatGPT in STEM and related subjects, language and writing skills, and skill training courses. In STEM subjects such as mathematics, science, and physics, researchers have explored how ChatGPT can assist students by explaining complex principles and concepts (Bitzenbauer, 2023; Leite, 2023) and providing problem-solving solutions (Sun et al. 2024). In language learning and writing courses, studies have explored the potential of ChatGPT for offering personalized suggestions for text revision (Polakova and Ivenz, 2024; Tsai et al. 2024), inspiring students' writing creativity (Nugroho et al. 2024), engaging students in spoken language practice (Young and Shishido, 2023), and helping students organize outlines for term papers (Su et al. 2023). In skill training courses, scholars have investigated ChatGPT's applications in simulating real-world scenarios (Lu et al. 2024) and collaborating with students to complete complex tasks (Darmawansah et al. 2024; Urban et al. 2024).

In terms of learning models, researchers have mainly explored the use of ChatGPT in supporting personalized learning, problem-based learning, and project-based learning. In personalized learning, studies have explored the potential of ChatGPT for providing targeted guidance to students (Almohesh, 2024; Gouia-Zarrad and Gunn, 2024), developing individualized learning plans (Lin, 2024), and assisting students in reviewing their assignments (Rahimi et al. 2025). In problem-based learning, scholars have investigated the use of ChatGPT for creating learning problem scenarios (Alsharif et al. 2024; Divito et al. 2024; Hui et al. 2025) and offering problem-solving solutions to students (Hamid et al. 2023; Urban et al. 2024). In project-based learning, studies have explored ChatGPT's applications in assisting students with designing, constructing, and completing projects (Avsheniuk et al. 2024; Economides and Perifanou, 2024; Villan and Santos, 2023).

With regard to role-playing, scholars have primarily explored the applications of ChatGPT as an intelligent tutor, an intelligent partner, and an intelligent learning tool. As an intelligent tutor, studies have examined the role ChatGPT can play in delivering knowledge to students (Ba et al. 2024; Ding et al. 2023) and assessing and grading student assignments (Boudouaia et al. 2024; Fokides and Peristeraki, 2025). As an intelligent partner, research to date has primarily focused on ChatGPT's application as a conversational practice robot (Xiao and Zhi, 2023; Young and Shishido, 2023). As an intelligent learning tool, scholars have investigated ChatGPT's role in assisting students with learning tasks (Su et al., 2023), quickly retrieving information (Pavlenko and Syzenko, 2024), and generating personalized learning content (Saleem et al. 2024).

Researchers have also explored the applications of ChatGPT in assessment, tutoring, and personalized recommendations. In terms of assessment, studies have examined the use of ChatGPT in grading student work (Chan et al. 2024; Su et al. 2023) and providing targeted, personalized feedback (Escalante et al. 2023; Qi et al. 2024). Regarding tutoring, scholars have explored the use of ChatGPT in addressing students' learning challenges during class (Huesca et al. 2024; Sun et al. 2024) and expanding their extracurricular knowledge (Ba et al. 2024). Regarding

personalized recommendations, studies have explored the role of ChatGPT in developing individualized learning plans for students (Albdrani and Al-Shargabi, 2023; Looi and Jia, 2025) and recommending diverse learning resources based on students' needs (Hartley et al. 2024; Jauhiainen and Guerra, 2024).

Empirical studies of ChatGPT on students' learning performance, learning perception, and higher-order thinking. Although many studies have examined the use of ChatGPT in the field of education, as of yet no consensus has been reached on whether it effectively supports students' learning performance, learning perception, and higher-order thinking. Some scholars argue for its benefits (Boudouaia et al. 2024; Gan et al. 2024; Zhou and Kim, 2024). For example, Emran et al. (2024) found that adopting ChatGPT as an assistive tool in an undergraduate academic writing skills course effectively improved the students' learning performance. Lu et al. (2024) used ChatGPT as an aid in a teacher training program and found that students aided by ChatGPT scored significantly higher on higher-order thinking than those in the control group who were only exposed to traditional teaching methods. Wu et al. (2023) conducted a quasi-experimental study in a high school mathematics course: the experimental group used ChatGPT for blended learning, while the control group used only iPads for blended learning. The results indicated that the experimental group significantly outperformed the control group in terms of intrinsic motivation, emotional engagement, and self-efficacy.

Conversely, other scholars have argued that ChatGPT hinders students' learning performance, learning perception, and higher-order thinking (Hays et al. 2024; Rahman and Watanobe, 2023). Niloy et al. (2023) conducted a quasi-experimental study with college students, in which the experimental group used ChatGPT 3.5 to assist with writing in the post-test, while the control group relied solely on publicly available secondary sources. Their results showed that the use of ChatGPT significantly reduced students' creative writing abilities. Yang et al. (2025) conducted a quasi-experimental study with high school students in a programming course. The experimental group used ChatGPT to assist with learning programming, while the control group was only exposed to traditional teaching methods. The results showed that the experimental group had lower flow experience, self-efficacy, and learning performance compared to the control group.

Finally, some experimental studies have demonstrated no significant difference between learning with ChatGPT and without it (Bašić et al. 2023a; Farah et al. 2023). Sun et al. (2024) used ChatGPT with a sample of students from a college programming course. Assessment of their performance revealed no significant difference in scores between the experimental and control groups. Donald et al. (2024) conducted a quasi-experimental study in a university programming course. Their results revealed no significant differences in programming performance or learning interest between the control group, which engaged in self-programming, and the experimental group, which used ChatGPT for assisted programming.

Researchers thus do not agree on whether ChatGPT is effective in promoting students' learning performance, learning perception, and higher-order thinking. As the research continues to deepen, researchers have also begun to explore the effectiveness of ChatGPT on student learning using literature reviews and meta-analyses.

Reviews and meta-analyses of ChatGPT and student learning. Most review studies have examined the effectiveness of ChatGPT in education (Ansari et al. 2024; Baig and Yadegaridehkordi, 2024; Chen et al. 2024; Grassini, 2023). Many of these reviews have reported both positive and negative impacts of ChatGPT on student learning. Regarding the positive impacts of ChatGPT, Ali et al. (2024) conducted a systematic review of 112 academic articles to explore the potential benefits and challenges of using ChatGPT in educational contexts. Their findings showed that ChatGPT can enhance student engagement, improve the accessibility of learning resources, strengthen language skills, and assist students in generating text. The study by Zhang and Tur (2024) focused on K-12 education to investigate the effects of ChatGPT on student learning. They found that ChatGPT provides learning support across different subjects, enhances personalized learning experiences, and improves learning efficiency. Moreover, Abu Khurma et al. (2024) conducted a systematic review of 16 related studies using the PRISMA method to analyze ChatGPT's role in promoting student engagement. They found that ChatGPT meets students' unique learning needs, fosters autonomy and motivation, and helps students gain a deeper understanding of content knowledge. Mai et al. (2024) conducted a SWOT analysis based on a systematic review and found that ChatGPT supports students in applying their knowledge to real-life situations and solving practical problems. Imran and Almusharraf (2023) explored ChatGPT's role in writing classes in higher education through a systematic review and concluded that ChatGPT improves writing efficiency, sparks creativity, aids translation, enhances content accuracy, and facilitates collaboration.

However, ChatGPT has also been found to have a negative impact on student learning. Ali et al. (2024) found that ChatGPT in education still has several shortcomings, such as generating incorrect answers, triggering academic plagiarism, and causing students to become dependent on technology. Zhang and Tur (2024) found that ChatGPT can reduce students' abilities, decrease teacher-student interaction, and mislead students with incorrect information. Abu Khurma et al. (2024) found that ChatGPT tends to make students overly reliant on technology, and can potentially provide inaccurate or biased information that can lead to cognitive biases and hinder the development of students' critical thinking skills. Dempere et al. (2023), through a systematic literature review, found that ChatGPT performs poorly in solving complex reasoning problems, fails to provide emotional support, and may negatively affect learning by reducing peer interaction. Lo et al. (2024) concluded that the use of ChatGPT can lead to plagiarism and academic misconduct, can provide incorrect information that can lower students' emotional attitudes toward learning, and can weaken students' independent thinking skills if they become excessively reliant on ChatGPT. Mai et al. (2024) argued that first-time users may find it challenging to identify inaccurate or misleading information generated by ChatGPT, adding to their cognitive load. Imran and Almusharraf (2023) pointed out that ChatGPT suffers from delayed information updates that may lead to harmful or incorrect content, which can mislead students and instill cognitive biases.

Although these literature reviews provide valuable insights into the application of ChatGPT in education, most of the studies they cover have been primarily theoretical, making it difficult to establish a causal relationship between ChatGPT use and student learning outcomes or to identify key moderating factors. This may explain the discrepancies in the conclusions drawn from different reviews. For example, Lo et al. (2024) argued that ChatGPT can either hinder or facilitate the development of students' critical thinking, while Abu Khurma et al. (2024) argued that ChatGPT only harms students' critical thinking. Therefore, it is essential to further analyze the actual impact of ChatGPT on student learning.

Some scholars have addressed this issue by conducting meta-analyses of experimental studies to search for the causal relationship between ChatGPT use and student learning. Heung

Table 1 Inclusion and exclusion criteria.

Inclusion criteria	• The research topics must be related to ChatGPT and learning performance, learning perception and higher-order thinking. • The research is written in English.
Exclusion criteria	• Not an experimental or quasi-experimental study. • Not experimental group with ChatGPT, control group without ChatGPT. • Does not contain data suitable for meta-analysis (e.g., Mean, SD, sample size, Cohen's *d*, *t* value).

and Chiu (2025), for example, conducted a meta-analysis to examine the impact of ChatGPT on student engagement. Their results indicate that ChatGPT can effectively enhance students' behavioral, cognitive, and emotional engagement in learning activities. Their study thus provides valuable insights into the impact of ChatGPT on student engagement. However, the scope of their study was somewhat narrow, as it focused solely on student engagement in learning and did not include any moderating variables, which limits their comprehensive analysis of the overall effectiveness of ChatGPT. Guan et al. (2024) focused on informal digital learning English (IDLE) and investigated the impact of ChatGPT on students' English proficiency, learning motivation, and self-regulation abilities. Their analysis revealed that ChatGPT use in IDLE contexts significantly contributes to enhancing English proficiency and self-regulation but has no significant effect on learning motivation. Again, their study makes a significant contribution to understanding the effectiveness of ChatGPT in the context of English learning. However, their meta-analysis was confined to a single domain, with a relatively narrow scope of focus. Deng et al. (2025) conducted a meta-analysis of 62 studies to explore the impact of ChatGPT on student learning. They found that ChatGPT can effectively enhance students' learning performance, improve the emotional-motivational aspects of learning, foster higher-order thinking, boost self-efficacy, and significantly reduce cognitive load. Their study greatly advances our understanding of the impact of ChatGPT on student learning, but there are some issues with the included samples. First, the samples selected for such a study should reflect students' actual learning after using ChatGPT as a learning aid. Simply using ChatGPT as a general tool without any learning process involved cannot adequately demonstrate its impact on student learning. For example, using ChatGPT solely for translation tasks (e.g., Moneus and Al-Wasy (2024) does not provide evidence for its influence on student learning. Additionally, ChatGPT itself does not directly affect student performance; the way in which the technology is applied is crucial to its effect. Therefore, both the experimental and control groups should adopt the same learning or teaching method to ensure comparability. However, in one of the sample studies (Beltozar-Clemente and Díaz-Vega, 2024) included by Deng et al. (2025), the experimental group received an intervention combining gamified learning with ChatGPT, but it was not specified whether the control group also used gamified learning. This omission may compromise the rigor of the study. While the findings of Deng et al. (2025) offer some reference value, their conclusions are insufficiently robust. Therefore, a more comprehensive meta-analysis is still needed to provide reliable results regarding the impacts of ChatGPT and the variables that modify its effects on students' learning performance, learning perception, and higher-order thinking.

Method

Research design. Examination of prior studies (Abuhassna and Alnawajha, 2023a, 2023b; Abuhassna et al. 2023; Masrom et al. 2021; Samsul et al. 2023) shows that most systematic literature reviews have been conducted following the preferred reporting items for systematic reviews and meta-analyses (PRISMA) guidelines, a well-known standard for systematic evaluation in many fields (Abuhassna and Alnawajha, 2023b). Therefore, this meta-analysis selected studies using the PRISMA guidelines and selection procedures (Moher et al. 2009; Page et al. 2021), as doing so can increase the transparency of the meta-analyses and ultimately improve the quality and reproducibility of the research (Page and Moher, 2017). The analysis in this study comprised four main steps: (1) conducting a literature search, (2) establishing inclusion and exclusion criteria, (3) coding study characteristics, and (4) performing data analysis.

Literature review process. The query search ("ChatGPT" OR "chat generative pre-trained transformer" OR "GPT-3.5" OR "GPT-4" OR "GPT-4o" OR "GenAI" OR "generative AI" OR "generative artificial intelligence" OR "Artificial Intelligence Generated Content" OR "AIGC") AND ("educat*" OR "learn*" OR "teach*") AND ("experiment*" OR "randomised controlled trial*" OR "randomised controlled trial*" OR "RCT*" OR "quasi-experiment*" OR "intervention") was conducted across multiple databases including the Web of Science Core Collection, Elsevier Science Direct, Taylor & Francis, Wiley Online Library, and Scopus. Since ChatGPT was launched by OpenAI in November 2022, the search period was restricted from November 2022 to February 2025. Initially, a total of 6,621 articles were obtained through database searches.

Inclusion and exclusion criteria. The literature inclusion and exclusion criteria were developed by two researchers specializing in educational technology, and are presented in Table 1.

After excluding duplicate articles ($n = 268$) and non-English publications ($n = 20$), samples were further screened by reviewing titles and abstracts to exclude studies unrelated to the impact of ChatGPT on student learning. Studies included in the review were required to investigate how ChatGPT affects student performance, learning perception, or higher-order thinking. Any study not meeting this criterion was excluded (e.g., research exploring factors influencing students' intention to use ChatGPT (Yu et al. 2024)). Studies that did not involve the learning process were also excluded (e.g., a study that merely used ChatGPT as a translation tool rather than as a way to enhance translation skills (Moneus and Al-Wasy, 2024)). At this stage, 3433 studies were removed. The meta-analysis of this study focuses specifically on ChatGPT. Thus, studies using other AI tools were excluded, such as Seth et al. (2024). This process resulted in the exclusion of an additional 126 studies, bringing the total to 3559 studies removed in this stage.

Subsequently, a thorough review of the remaining 2712 articles was conducted. First, studies that did not use quasi-experimental or experimental designs were excluded. To be included, experimental studies had to include at least one experimental group and one control group and employ methods such as random assignment, statistical matching, or controlling for pretest differences. Therefore, studies with non-experimental or non-quasi-experimental designs were excluded, such as Jošt et al. (2024). At this stage, 1455 articles were removed.

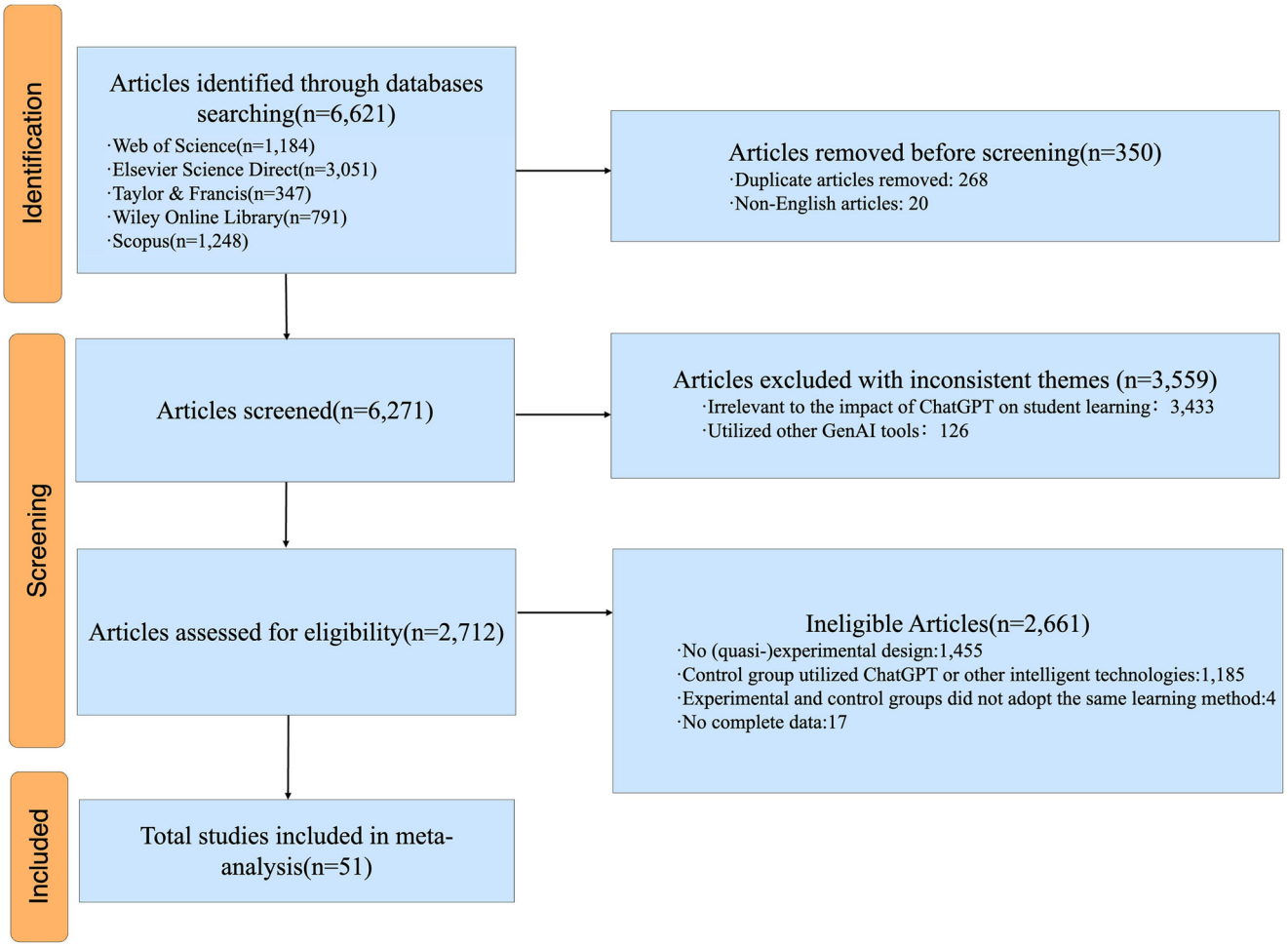

Fig. 1 PRISMA.

Given that the purpose of this study is to examine whether ChatGPT effectively promotes student learning, the included studies had to strictly meet the criterion by which the experimental group used ChatGPT, while the control group did not. Studies that did not adhere to this requirement were excluded (e.g., Ng et al. 2024). To ensure the rigor of the research findings, permissible interventions for the control group were limited to traditional tools such as pen and paper, PowerPoint, and essential learning materials required for the course (e.g., computers in programming courses). Studies whose control groups used emerging technologies with intelligent features, such as augmented reality (AR), virtual reality (VR), metaverse, digital twins, holographic displays, or domain-specific intelligent tools, were also excluded (e.g., Coban et al. 2025; W. S. Wang et al. 2024a). At this stage, 1185 studies were removed. Furthermore, the experimental and control groups had to adopt the same learning method. Studies that did not clearly specify that both groups used the same learning method were excluded (e.g., Beltozar-Clemente and Díaz-Vega 2024). At this stage, 4 articles were removed.

This study used comprehensive meta-analysis software (version 3.3.0) to calculate the standardized mean difference Hedges's *g* for the effect of ChatGPT on student learning performance, learning perception, and higher-order thinking. The calculation was based on either: (a) the sample sizes, means, and standard deviations of the post-test data of the experimental and control groups, or (b) the sample sizes, means, and standard deviations of the pre-test and post-test data from a single experiment. When means or standard deviations were unavailable, χ^2, *t*, and *F* values were used instead. The included samples were thus required to report at least the sample size, mean, and standard deviation, or else the χ^2, *t*, and *F* values. Samples with insufficient data to compute Hedges's *g* were excluded as well. At this stage, 17 articles were removed.

Articles that met the coding criteria were independently selected by two researchers, and acceptable interrater reliability was achieved (Cronbach's $\alpha = 0.96$). Any disagreements were resolved through consensus reached via repeated discussion between the researchers. The final sample consisted of 51 articles, published between November 2022 and February 2025. Figure 1 illustrates the selection process (see Supplementary Fig. 1 PRISMA).

Coding of study characteristics. To analyze the impact of ChatGPT on student learning performance, learning perception, and higher-order thinking, a specific coding scheme was needed for a comprehensive analysis of the 51 selected articles. This scheme was meticulously developed by thoroughly examining the included literature and drawing on insights by Zheng et al. (2021). A total of six moderating variables were selected (see Table 2).

The coding of moderating variables followed the scheme outlined in Table 2. The data within the articles were coded

Table 2 Coding scheme for moderating variables.

Variables	Category
Grade level	1. Primary (ages 7–13) 2. Secondary (ages 14–17) 3. College (ages over 18)
Type of course	1. STEM and related courses (programming, mathematics, science, physics, etc.) 2. Language learning and academic writing (English, academic writing courses, research methodology courses, etc.) 3. Skills and competencies development (teacher training courses, developing creative problem-solving skills courses, etc)
Learning model	1. Personalized learning (An instructional approach that tailors learning experiences to individual student's needs, preferences, and pace, as exemplified by Song and Song (2023)) 2. Problem-based learning (A student-centered method where learners develop knowledge and skills by solving authentic, real-world problems, as demonstrated by Urban et al. (2024)) 3. Project-based learning (An active learning approach in which students engage in extended projects to explore complex topics, as illustrated in Küchemann et al. (2023)) 4. Contextual learning (A learning approach that situates knowledge within specific, real-life contexts, as shown in Lu et al. (2024)) 5. Reflective learning (A process of reviewing, analyzing, and evaluating one's learning process or outcomes, followed by necessary adjustments, as exemplified by Escalante et al. (2023)) 6. Mixed
Duration	1. ≤1 week 2. 1–4 weeks 3. 4–8 weeks 4. >8 weeks
Role of ChatGPT	1. Intelligent tutor (primarily functions as a teacher, providing personalized guidance, feedback, and assessment to students, as exemplified by Escalante et al. (2023)) 2. Intelligent partner (acts as a peer, engaging in dialogue and interaction with students to simulate the behavior of classmates, as shown in Li et al. (2024)) 3. Intelligent learning tool (focuses more on offering auxiliary functions, such as resource recommendations, knowledge retrieval, and learning progress tracking, without simulating human interaction, as exemplified by Song and Song (2023)) 4. Mixed
Area of ChatGPT application	1. Assessment and evaluation (using ChatGPT to provide assessments and feedback on student learning, as exemplified by Escalante et al. (2023)) 2. Tutoring (mainly refers to students using ChatGPT on their own during or after class to solve difficulties encountered in learning, as shown in Essel et al. (2024)) 3. Personalized recommendation (mainly refers to ChatGPT providing corresponding learning resources and suggestions based on student's individual needs, as exemplified by Song and Song (2023)) 4. Mixed
Learning Effects	1. Higher-order thinking (primarily measured through scales and questionnaires, as shown in Lu et al. (2024)) 2. Learning perception (primarily assessed through scales and questionnaires, as shown in Chen and Chang (2024)) 3. Learning performance (primarily evaluated through tests and exams, as demonstrated in Urban et al. (2024))

according to the following criteria: (a) The effect sizes for each independent sample in the study were recorded. (b) When a study included two or more experiments, each experiment was counted as one effect size. (c) When a study reported data on different dimensions of the same variable (e.g., dividing higher-order thinking into critical thinking and creative thinking, and reporting separate scores for each), the average effect size across all dimensions was coded. Ultimately, the coding process resulted in the extraction of 72 independent effect sizes from the selected 51 articles.

Data analysis. Comprehensive Meta-Analysis 3.0 (CMA3.0) software was used to process the collected data. Following the example of Lipsey and Wilson (2001), effect sizes and variances were extracted from each study. The scores of two or more groups for each condition were averaged before calculating the effect size. Most of the included studies reported the experimental and control groups' post-test means and standard deviations, and these were used to calculate the standardized mean difference effect sizes. Since differences in sample sizes between studies might introduce bias into the calculated effect sizes, Hedges's g was used to adjust the effect sizes to correct for small sample size biases (Hedges, 1981). The calculated effect sizes were interpreted according to guidelines proposed by Sawilowsky (2009): very small (0.1), small (0.2), medium (0.5), large (0.8), very large (1.2), and huge (2.0).

Results

Table 3 presents the 72 effects obtained from the coding of the 51 studies, including the sample size (N), independent effect size (g), grade level, type of course, duration, learning model, role of ChatGPT, area of ChatGPT application, and learning effects.

Heterogeneity. This study used the Q-test and I^2 to calculate the level of heterogeneity within the sample. Cochran's Q mainly examines the p-value, with a value less than 0.1 indicating heterogeneity and that the random effects model should be used for meta-analysis (Zhang et al., 2024). I^2 assesses the degree of heterogeneity, with 0–25% indicating low heterogeneity, 25–75% indicating moderate heterogeneity, and 75–100% indicating high heterogeneity (Higgins et al., 2003). According to Table 4, the Q-value for learning performance is 339.725, and the p-value of the Q-test is <0.1. The I^2 value is 89.243%, indicating a high degree of heterogeneity in the included studies regarding the impact of ChatGPT on learning performance. The Q-value for learning perception is 65.646, and the p-value of the Q-test is less than 0.1. The I^2 value is 72.580%, indicating a moderate degree of heterogeneity in the included studies regarding the effect of

Table 3 Studies included in the meta-analysis.

Learning effect	Author(year)	*N*		*g*	Grade level	Type of course	Duration	Learning model	Role of ChatGPT	Area of ChatGPT application
		EG	CG							
Learning performance	Emran et al. (2024)	34	34	2.730	College	Language learning and academic writing	>8 week	Problem-based learning	Intelligent learning tool	Tutoring
Learning performance	Almohesh (2024)	75	75	1.117	Primary	Skills and competencies development	≤1 week	Personalized learning	Intelligent tutor	Mixed
Learning performance	Avello-Martínez et al. (2023)	20	21	0.062	College	Skills and competencies development	≤1 week	Personalized learning	Intelligent learning tool	Personalized recommendation
Learning performance	Boudouaia et al. (2024)	37	39	0.797	College	Language learning and academic writing	>8 week	Reflective learning	Intelligent tutor	Assessment and evaluation
Learning performance	Bašić et al. (2023a)	12	12	0.993	College	STEM and related courses	≤1 week	Problem-based learning	Intelligent learning tool	Tutoring
Learning performance	Chen and Chang (2024)	31	30	1.235	College	STEM and related courses	1–4 week	Mixed	Intelligent learning tool	Mixed
Learning performance	Chen and Chang (2024)	63	70	0.169	Secondary	STEM and related courses	1–4 week	Mixed	Intelligent learning tool	Tutoring
Learning performance	Song and Song (2023)	25	25	0.790	College	Language learning and academic writing	>8 week	Mixed	Intelligent learning tool	Mixed
Learning performance	Sun et al. (2024)	43	39	0.280	College	STEM and related courses	4–8 week	Mixed	Intelligent learning tool	Tutoring
Learning performance	Darmawansah et al. (2024)	34	33	0.697	College	Skills and competencies development	4–8 week	Mixed	Mixed	Mixed
Learning performance	Donald et al. (2024)	16	17	−0.334	College	STEM and related courses	1–4 week	Personalized learning	Mixed	Mixed
Learning performance	Li (2023)	42	39	0.493	College	STEM and related courses	4–8 week	Mixed	Intelligent tutor	Mixed
Learning performance	Ba et al. (2024)	39	38	0.067	College	Skills and competencies development	1–4 week	Contextual learning	Intelligent tutor	Tutoring
Learning performance	Hsu (2024)	20	20	2.420	College	Skills and competencies development	>8 week	Personalized learning	Mixed	Mixed
Learning performance	Hu (2024)	67	68	1.8985	College	Skills and competencies development	4–8 week	Problem-based learning	Intelligent partner	Mixed

Table 3 (continued)

Learning effect	Author(year)	*N*		*g*	Grade level	Type of course	Duration	Learning model	Role of ChatGPT	Area of ChatGPT application
		EG	CG							
Learning performance	Huang et al. (2024)	60	59	0.162	College	STEM and related courses	≤1 week	Mixed	Intelligent tutor	Personalized recommendation
Learning performance	Huesca et al. (2024)	214	128	0.453	College	STEM and related courses	>8 week	Personalized learning	Intelligent learning tool	Tutoring
Learning performance	Roganović (2024)	16	44	0.037	College	STEM and related courses	≤1 week	Problem-based learning	Intelligent learning tool	Tutoring
Learning performance	Lu et al. (2024)	107	108	0.683	College	Skills and competencies development	1–4 week	Contextual learning	Mixed	Mixed
Learning performance	Escalante et al. (2023)	23	25	0.500	College	Language learning and academic writing	4–8 week	Reflective learning	Intelligent tutor	Assessment and evaluation
Learning performance	Mahapatra (2024)	35	37	1.175	College	Language learning and academic writing	4–8 week	Personalized learning	Mixed	Mixed
Learning performance	Urban et al. (2024)	77	68	0.612	College	Skills and competencies development	≤1 week	Problem-based learning	Intelligent learning tool	Mixed
Learning performance		20	20	2.420	College	STEM and related courses	>8 week	Problem-based learning	Intelligent learning tool	Tutoring
Learning performance	Ghafouri, et al. (2024)	24	24	1.746	College	Language learning and academic writing	>8 week	Problem-based learning	Intelligent learning tool	Assessment and evaluation
Learning performance	Mugableh (2024)	30	30	1.553	College	Language learning and academic writing	>8 week	Personalized learning	Intelligent tutor	Assessment and evaluation
Learning performance	Karaman and Goksu (2024)	24	15	0.070	Secondary	STEM and related courses	4–8 week	Mixed	Intelligent learning tool	Tutoring
Learning performance	Jafarian et al. (2025)	120	153	0.047	College	STEM and related courses	>8 week	Personalized learning	Intelligent learning tool	Tutoring
Learning performance	Pan et al. (2025)	30	31	0.798	College	Language learning and academic writing	>8 week	Personalized learning	Intelligent learning tool	Personalized recommendation

Table 3 (continued)

Learning effect	Author(year)	N		g	Grade level	Type of course	Duration	Learning model	Role of ChatGPT	Area of ChatGPT application
		EG	CG							
Learning performance	Qi et al. (2024)	23	23	1.105	College	Language learning and academic writing	4–8 week	Problem-based learning	Intelligent tutor	Assessment and evaluation
Learning performance	Alneyadi and Wardat (2023)	58	64	0.973	Secondary	STEM and related courses	>8 week	Problem-based learning	Intelligent learning tool	Tutoring
Learning performance	Rakap (2023)	11	11	2.857	College	Skills and competencies development	≤1 week	Reflective learning	Intelligent learning tool	Tutoring
Learning performance	An et al. (2025)	55	55	1.374	College	Skills and competencies development	4–8 week	Project-based learning	Mixed	Mixed
Learning performance	Küchemann et al. (2023)	13	13	0.688	Secondary	STEM and related courses	≤1 week	Problem-based learning	Intelligent learning tool	Tutoring
Learning performance	Chan et al. (2024)	342	576	0.195	College	Language learning and academic writing	≤1 week	Project-based learning	Intelligent tutor	Assessment and evaluation
Learning performance	Wu et al. (2024)	11	12	1.537	College	STEM and related courses	4–8 week	Project-based learning	Intelligent partner	Tutoring
Learning performance	Kosar et al. (2024)	99	83	−0.034	College	STEM and related courses	>8 week	Mixed	Intelligent learning tool	Mixed
Learning performance	Yang et al. (2025)	76	77	0.315	Secondary	STEM and related courses	1–4 week	Personalized learning	Mixed	Mixed
Learning performance	Wang et al. (2024a, 2024b)	36	38	0.712	College	Skills and competencies development	>8 week	Personalized learning	Intelligent learning tool	Assessment and evaluation
Learning performance	Gan et al. (2024)	55	55	0.281	College	STEM and related courses	1–4 week	Personalized learning	Intelligent learning tool	Tutoring
Learning performance	Ji et al. (2023)	21	15	4.009	College	STEM and related courses	4–8 week	Mixed	Intelligent partner	Tutoring
Learning performance	Basic et al. (2023b)	9	9	−0.516	College	Language learning and academic writing	≤1 week	Problem-based learning	Intelligent learning tool	Tutoring
Learning performance	Hui et al. (2025)	21	21	1.146	College	Skills and competencies development	1–4 week	Problem-based learning	Intelligent partner	Mixed
Learning perception	Chen and Chang (2024)	31	30	1.290	College	STEM and related courses	1–4 week	Mixed	Intelligent learning tool	Mixed

Table 3 (continued)

Learning effect	Author(year)	*N*		*g*	Grade level	Type of course	Duration	Learning model	Role of ChatGPT	Area of ChatGPT application
		EG	CG							
Learning perception	Chen and Chang (2024)	63	70	−0.059	Secondary	STEM and related courses	1-4 week	Problem-based learning	Intelligent learning tool	Tutoring
Learning perception	Song and Song (2023)	25	25	0.626	College	Language learning and academic writing	>8 week	Personalized learning	Intelligent learning tool	Mixed
Learning perception	Ng et al. (2024)	37	37	0.864	Secondary	STEM and related courses	1-4 week	Mixed	Intelligent tutor	Tutoring
Learning perception	Donald et al. (2024)	16	17	−0.457	College	STEM and related courses	1-4 week	Personalized learning	Mixed	Mixed
Learning perception	Li (2023)	42	39	0.476	College	STEM and related courses	4-8 week	Mixed	Intelligent tutor	Mixed
Learning perception	Huang et al. (2024)	60	59	0.180	College	STEM and related courses	≤1 week	Mixed	Intelligent tutor	Personalized recommendation
Learning perception	Ji et al. (2023)	21	15	0.342	College	STEM and related courses	4-8 week	Project-based learning	Intelligent learning tool	Mixed
Learning perception	Lu et al. (2024)	107	108	0.403	College	Skills and competencies development	1-4 week	Contextual learning	Mixed	Mixed
Learning perception	Liao et al. (2024)	26	26	−0.007	College	STEM and related courses	4-8 week	Project-based learning	Intelligent learning tool	Tutoring
Learning perception	Urban et al. (2024)	77	68	0.337	College	Skills and competencies development	≤1 week	Problem-based learning	Intelligent learning tool	Mixed
Learning perception	Ghafouri, et al. (2024)	24	24	1.622	College	Language learning and academic writing	>8 week	Problem-based learning	Intelligent learning tool	Assessment and evaluation
Learning perception	Shang et al. (2024)	55	54	0.320	College	STEM and related courses	≤1 week	Problem-based learning	Intelligent learning tool	Tutoring
Learning perception	Chan et al. (2024)	342	576	0.180	College	Language learning and academic writing	≤1 week	Project-based learning	Intelligent tutor	Assessment and evaluation
Learning perception	Wu et al. (2024)	11	12	0.876	College	STEM and related courses	4-8 week	Project-based learning	Intelligent tutor	Tutoring
Learning perception	Yang et al. (2025)	76	77	0.486	Secondary	STEM and related courses	1-4 week	Personalized learning	Mixed	Mixed
Learning perception	Wu et al. (2024)	35	35	0.474	Secondary	STEM and related courses	1-4 week	Personalized learning	Intelligent learning tool	Tutoring

Table 3 (continued)

Learning effect	Author(year)	*N*		*g*	Grade level	Type of course	Duration	Learning model	Role of ChatGPT	Area of ChatGPT application
		EG	CG							
Learning perception	Yilmaz et al. (2023)	21	24	0.296	College	STEM and related courses	4–8 week	Personalized learning	Intelligent learning tool	Tutoring
Learning perception	Ji et al. (2023)	21	15	1.533	College	STEM and related courses	4–8 week	Mixed	Intelligent tutor	Tutoring
High order thinking	Ng et al. 2024	37	37	0.943	Secondary	STEM and related courses	1–4 week	Mixed	Intelligent tutor	Tutoring
High order thinking	Li (2023)	42	39	0.541	College	STEM and related courses	4–8 week	Mixed	Intelligent learning tool	Mixed
High order thinking	Essel et al. (2024)	60	65	0.157	College	Language learning and academic writing	>8 week	Project-based learning	Intelligent learning tool	Tutoring
High order thinking	Lu et al. (2024)	107	108	0.039	College	Skills and competencies development	1–4 week	Contextual learning	Mixed	Mixed
High order thinking	Kareem et al. (2025)	150	150	0.448	College	Skills and competencies development	1–4 week	Project-based learning	Intelligent learning tool	Tutoring
High order thinking	Urban et al. (2024)	77	68	0.366	College	Skills and competencies development	≤1 week	Problem-based learning	Intelligent learning tool	Mixed
High order thinking	Pan et al. (2025)	30	31	0.706	College	Language learning and academic writing	>8 week	Personalized learning	Intelligent learning tool	Personalized recommendation
High order thinking	Wu et al. (2024)	11	12	0.950	College	STEM and related courses	4–8 week	Project-based learning	Intelligent tutor	Tutoring
High order thinking	Yilmaz et al. (2023)	21	24	0.710	College	STEM and related courses	4–8 week	Personalized learning	Intelligent learning tool	Tutoring

EG Experimental Group, *CG* Control Group.

Table 4 Results of overall effects analysis of ChatGPT to students' learning performance, learning perception, and higher-order thinking.

Learning effect	k	g	95%CI		Two-tailed test		Test of homogeneity			
			LL	UL	Z	P	Q	df(Q)	I^2	p
Learning performance	44	0.867	0.683	1.052	8.458	<0.001	399.725	43	89.243	<0.001
Learning perception	19	0.456	0.287	0.625	5.289	<0.001	65.646	18	72.580	<0.001
Higher-order thinking	9	0.457	0.255	0.650	4.471	<0.001	18.105	8	55.813	<0.05

k = number of independent studies, g = mean effect size, CI = confidence interval.

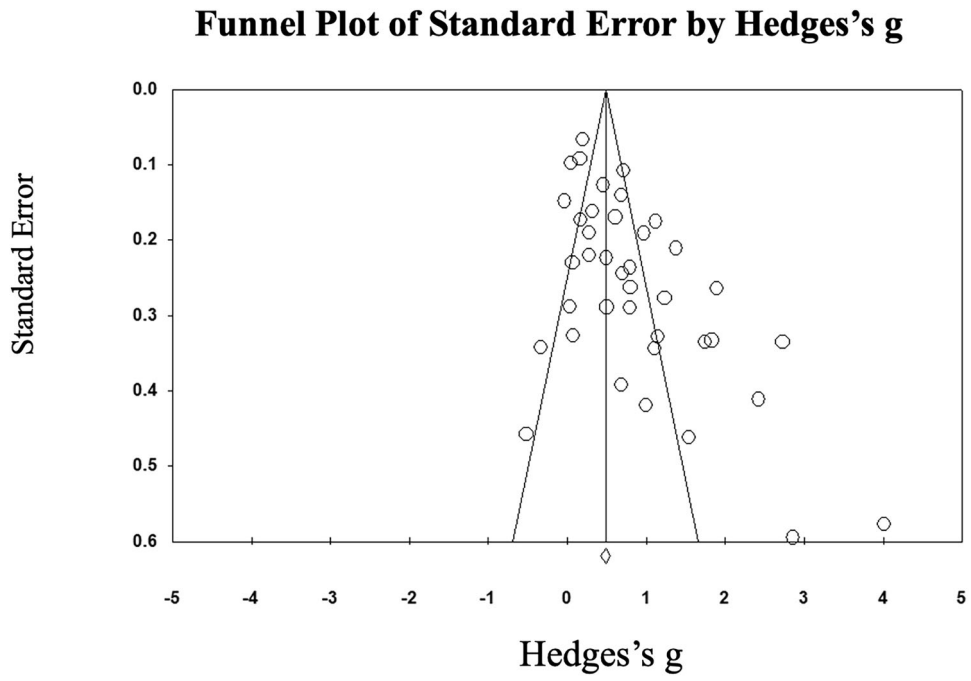

Fig. 2 Funnel plot for learning performance.

ChatGPT on learning perception. The Q-test for higher-order thinking is 18.105, and the p-value of the Q-test is less than 0.1. The I^2 value is 55.813%, indicating moderate heterogeneity in the included studies regarding the effect of ChatGPT on higher-order thinking. Consequently, the random effects model was selected for the analysis of learning performance, learning perception, and higher-order thinking.

Publication bias and sensitivity analysis. The funnel plot and fail-safe N test were used to assess potential publication bias. In the absence of publication bias, a funnel plot should resemble a symmetrical inverted funnel, whereas an asymmetrical or incomplete funnel indicates the possibility of bias (Schulz et al. 1995). The funnel plots for the present research are shown in Supplementary Figs. 2 through 4 (see Supplementary Fig. 2 Funnel plot for learning performance, Supplementary Fig. 3 Funnel plot for learning perception, and Supplementary Fig. 4 Funnel plot for higher-order thinking). The effect values for the impact of ChatGPT on learning performance, learning perception, and higher-order thinking are relatively evenly distributed on both sides of the funnel plot, indicating that the data have sufficiently high reliability to be used for meta-analysis. However, some points are outside the two diagonal lines, indicating potential bias or heterogeneity in the included studies, possibly due to differences in sample size or research design (Figs. 2–4).

Meanwhile, if the calculated fail-safe N is greater than the tolerance value of $5k + 10$, where k is the sample size, the estimated effect size of an unpublished study is considered unlikely to influence the overall effect size of the meta-analysis (Borenstein et al. 2009). In this study, the fail-safe N for learning performance was calculated to be 5490 ($5k + 10 = 230$, $k = 44$), for learning perception was calculated to be 399 ($5k + 10 = 105$, $k = 19$), and for higher-order thinking was calculated to be 101 ($5k + 10 = 55$, $k = 9$), indicating that the possibility of publication bias is relatively small for this study.

Because some points lay outside the two diagonal lines, a sensitivity analysis was conducted to further assess the robustness of the meta-analysis results: this technique is used to examine outliers that might affect the magnitude of the overall effect. One-Study-Removal Analysis was used to detect the impact of extreme positive and negative effects on the overall effect. The results show that, for learning performance, after the removal of any single study, the range of the 95% confidence interval (CI) of the effect size still remained between 0.462 and 0.574 for the fixed-effects model and between 0.683 and 1.052 for the random-effects model. For learning perception, after the removal of any single study, the range of the 95% CI of the effect size remained between 0.235 and 0.387 for the fixed-effects model and between 0.287 and 0.625 for the random-effects model. For higher-order thinking, after the removal of any single study, the range of the 95% CI of the effect size remained between 0.266 and 0.509 for the fixed-

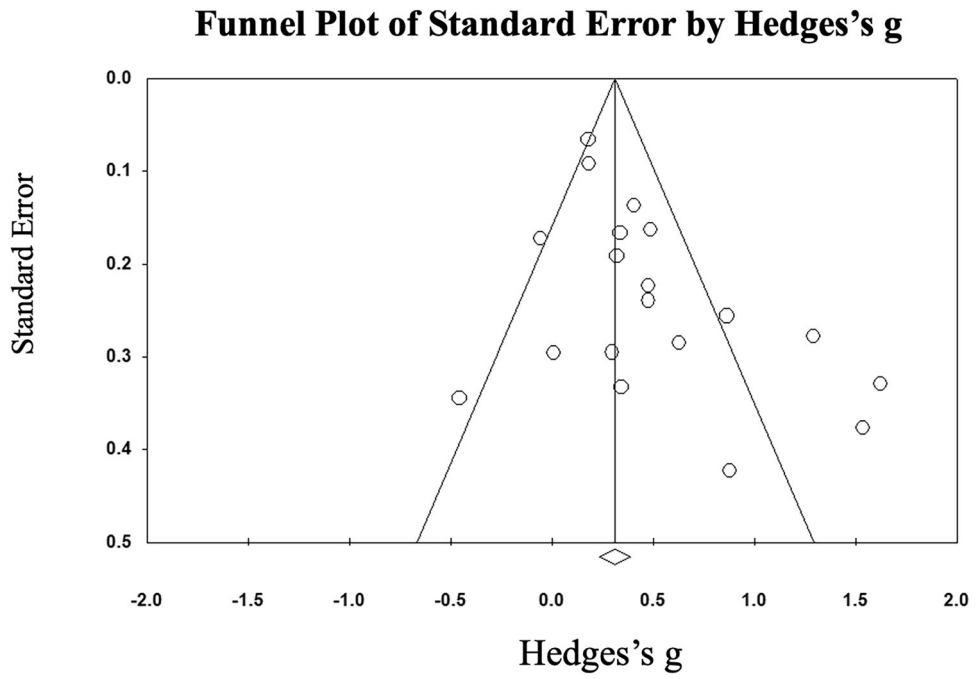

Fig. 3 Funnel plot for learning perception.

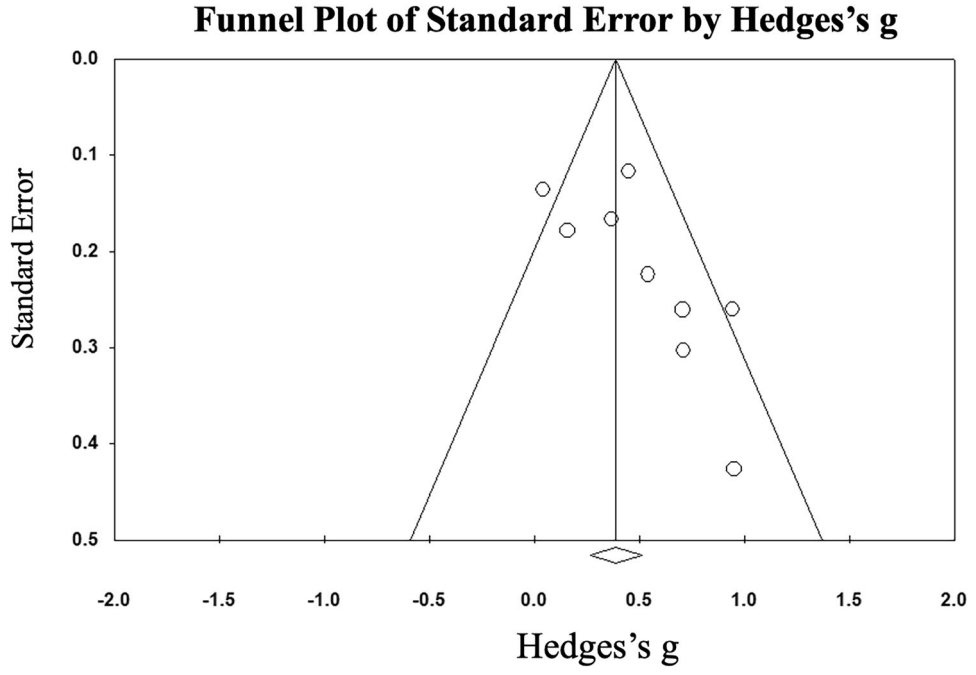

Fig. 4 Funnel plot for higher-order thinking.

effects model and between 0.255 and 0.650 for the random-effects model. Thus, the meta-analysis results obtained in this study were demonstrated to be very stable.

Overall effectiveness. The results of the meta-analysis are presented in Table 4. The overall effects of ChatGPT on enhancing learning performance, improving learning perception, and promoting higher-order thinking were calculated as $g = 0.867$, $g = 0.456$, and $g = 0.457$, respectively, indicating that ChatGPT has a large positive effect on learning performance and a medium positive effect on learning perception and higher-order thinking.

The effects of the moderating variables. Separate moderator analyses were conducted for learning performance, learning perception, and higher-order thinking to explore possible causes of heterogeneity.

Table 5 presents the results of the moderator analysis for learning performance. Significant differences were observed among the types of courses ($Q_B = 64.249$, $p < 0.001$), with skills and competencies development ($g = 0.874$) showing the strongest effect. Learning models also exhibited significant differences ($Q_B = 76.220$ $p < 0.001$), primarily because ChatGPT had a larger effect in problem-based learning ($g = 1.113$), than in other learning models. Similarly, duration showed significant

Table 5 Results of moderating variables analysis of ChatGPT to learning performance.

Moderator	k	g(95%CI)	Z	df(Q)	Q_B	P-value
Grade level					0.627	0.428
Secondary	5	0.431[0.248, 0.614]	4.624***	4		
College	38	0.509[0.449, 0.568]	16.773***	37		
Type of courses					64.249***	<0.001
STEM and related courses	19	0.311[0.229,0.394]	7.418***	18		
Skills and competencies development	13	0.874[0.764,0.985]	15.069***	12		
Language learning and academic writing	12	0.531[0.310,0.531]	7.454***	11		
Learning model					76.220***	<0.001
Contextual learning	2	0.517[0.283,0.752]	6.509***	1		
Personalized learning	13	0.535[0.440,0.630]	6.654***	12		
Problem-based learning	12	1.113[0.957,1.270]	13.914***	11		
Project-based learning	3	0.239[0.112,0.367]	3.677***	2		
Reflective learning	3	0.866[0.524,1.209]	4.956***	2		
Mixed	11	0.330[0.217,0.443]	5.200***	10		
Duration					55.998***	<0.001
≤1 week	11	0.332[0.242,0.423]	11.912***	10		
1–4 week	8	0.428[0.289,0.566]	6.051***	7		
4–8 week	12	0.999[0.843,1.156]	12.525***	11		
>8 week	13	0.531[0.490,0.684]	11.912***	12		
Role of ChatGPT					6.261	0.100
Intelligent tutor	11	0.558[0.454,0.662]	9.722***	10		
Intelligent partner	3	0.399[0.235,0.564]	4.769***	2		
Intelligent learning tool	21	0.480[0.397,0.564]	9.203***	20		
Mixed	9	0.645[0.499,0.791]	6.747***	8		
Area of ChatGPT Application					6.210	0.102
Assessment and evaluation	6	0.418[0.315,0.521]	7.943***	5		
Personalized recommendation	4	0.687[0.502,0.872]	7.287***	3		
Tutoring	17	0.480[0.377,0.583]	9.134***	16		
Mixed	12	0.485[0.380,0.591]	7.961***	11		

k = number of independent studies, g = mean effect size, CI = confidence interval, Q_B = between-group homogeneity.
*$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$.

differences ($Q_B = 55.998$, $p < 0.001$), with ChatGPT demonstrating a large effect in interventions lasting 4–8 weeks ($g = 0.999$).

No significant differences were found for grade level ($Q_B = 0.627$ $p = 0.428$), roles of ChatGPT ($Q_B = 6.261$, $p = 0.100$), and area of ChatGPT application ($Q_B = 6.210$, $p = 0.102$). It should be noted that the primary school sample was excluded from the analysis because it consisted of only one study, which did not find a statistically significant effect.

Table 6 presents the results of the moderator analysis for learning perception. Of all the tested moderating variables, intervention duration was the only one to have a significant effect ($Q_B = 19.839$, $p < 0.001$). ChatGPT was found to be beneficial across all time periods, with interventions lasting more than 8 weeks ($g = 1.054$) showing a substantially larger effect size than those lasting less than 1 week ($g = 0.203$), 1–4 weeks ($g = 0.404$), and 4–8 weeks ($g = 0.493$). No significant differences were observed for grade level ($Q_B = 0.360$, $p = 0.562$), type of courses ($Q_B = 1.338$, $p = 0.512$), learning models ($Q_B = 5.752$, $p = 0.124$), role of ChatGPT ($Q_B = 2.748$, $p = 0.253$), or area of ChatGPT application ($Q_B = 4.861$, $p = 0.088$).

Table 7 presents the results of the moderator analysis for higher-order thinking. All relevant studies were conducted at the college level, and thus no analysis was performed for grade level. Significant differences were observed for type of courses ($Q_B = 7.811$, $p < 0.05$). Specifically, ChatGPT had a large positive effect on higher-order thinking for STEM and related courses ($g = 0.737$), a nearly medium positive effect for language learning and academic writing ($g = 0.334$), and a small positive effect for skills and competencies development ($g = 0.296$).

There were significant differences for roles of ChatGPT as well ($Q_B = 4.872$, $p < 0.05$). ChatGPT was most effective when acting as an intelligent tutor ($g = 0.945$), having the largest impact on higher-order thinking. Mixed roles and intelligent partners were excluded from the analysis due to a lack of sufficient data, as each was represented by only one study.

No significant moderating effects were detected for the learning model ($Q_B = 4.109$, $p = 0.128$), duration ($Q_B = 2.988$, $p = 0.224$), or application area ($Q_B = 5.073$, $p = 0.079$).

Discussion

How effective is ChatGPT in promoting learning performance?

The results indicate that ChatGPT has a substantial positive effect on students' learning performance, with an overall effect size of 0.867. The meta-analysis further revealed that the impact of ChatGPT on students' learning performance is moderated by three variables: type of course, learning models, and intervention duration. The overall effect size indicated that ChatGPT is effective in improving students' learning performance, a result that is consistent not only with prior meta-analyses (Deng et al. 2025) but also with the findings of most empirical studies (Alneyadi and Wardat, 2023; An et al. 2025; Darmawansah et al. 2024). Deng et al. (2025) argued that ChatGPT may genuinely enhance learning performance by enabling personalized learning experiences, providing immediate access to information and diverse perspectives, and allowing students to engage more deeply with the material. A quasi-experimental study by Alneyadi and Wardat (2023) showed that students who incorporated ChatGPT into their studies tended to acquire additional knowledge, thereby improving their understanding of complex concepts and enhancing their learning performance. Interestingly, the current study compares the results of the present meta-analysis with those of

Table 6 Results of moderating variables analysis of ChatGPT to learning perception.

Moderator	*k*	g(95%CI)	*Z*	df(Q)	Q_B	*P*-value
Grade level					0.360	0.562
Secondary	4	0.363[0.171, 0.556]	3.700***	3		
College	15	0.301[0.219, 0.384]	7.194***	14		
Type of courses					1.338	0.512
STEM and related courses	14	0.335[0.228,0.442]	6.134***	13		
Skills and competencies development	2	0.376[0.168,0.584]	3.548***	1		
Language learning and academic writing	3	0.255[0.131,0.380]	4.029***	2		
Learning model					5.752	0.124
Personalized learning	5	0.387[0.177,0.597]	3.613***	4		
Problem-based learning	4	0.319[0.129,0.510]	3.287***	3		
Project-based learning	4	0.193[0.070,0.316]	3.074***	3		
Mixed	5	0.412[0.263,0.560]	5.425***	4		
Duration					19.839***	<0.001
≤1 week	4	0.203[0.106,0.300]	4.105***	3		
1–4 week	6	0.404[0.228,0.579]	4.512***	5		
4–8 week	6	0.493[0.249,0.737]	3.959***	5		
>8 week	2	1.054[0.632,1.477]	4.893***	1		
Role of ChatGPT					2.748	0.253
Intelligent tutor	6	0.256[0.157,0.354]	5.072***	5		
Intelligent learning tool	10	0.405[0.259,0.550]	5.432***	9		
Mixed	2	0.313[0.024,0.602]	2.122***	1		
Area of ChatGPT application					4.861	0.088
Assessment and evaluation	2	0.236[0.109,0.364]	3.632***	1		
Tutoring	8	0.367[0.194,0.540]	4.167***	7		
Mixed	8	0.448[0.306,0.591]	6.166***	7		

k = number of independent studies, *g* = mean effect size, CI = confidence interval, Q_B = between-group homogeneity,
*$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$.

Table 7 Results of moderating variables analysis of ChatGPT to higher-order thinking.

Moderator	*k*	g(95%CI)	*Z*	df(Q)	Q_B	*P*-value
Type of courses					7.811*	<0.05
STEM and related courses	4	0.737[0.463,1.012]	5.269***	3		
Skills and competencies development	3	0.295[0.141,0.448]	3.767***	2		
Language learning and academic writing	2	0.332[0.043,0.624]	1.687	1		
Learning model					4.109	0.128
Personalized learning	2	0.718[0.325,1.111]	4.193***	1		
Project-based learning	3	0.391[0.204,0.579]	4.091***	2		
Mixed	2	0.719[0.383,1.056]	4.193***	1		
Duration					2.988	0.224
1–4 week	3	0.344[0.180,0.508]	4.103***	2		
4–8 week	3	0.654[0.329,0.979]	3.939***	2		
>8 week	2	0.332[0.043,0.620]	2.254**	1		
Role of ChatGPT					4.872*	<0.05
Intelligent tutor	2	0.945[0.509,1.380]	4.254***	1		
Intelligent learning tool	6	0.428[0.285,0.572]	5.850***	5		
Area of ChatGPT application					5.073	0.079
Tutoring	5	0.478[0.308,0.647]	2.129***	4		
Mixed	3	0.238[0.050,0.425]	5.536***	2		

k = number of independent studies, *g* = mean effect size, CI = confidence interval, Q_B = between-group homogeneity.
*$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$.

another meta-analysis that focused on the effects of AI-based assessment tools on learners' learning performance (Chen et al. 2025). The results indicate that the effect size for the impact of ChatGPT on students' learning performance is significantly larger than that of other AI assessment tools ($g = 0.390$). This may be because the capabilities of traditional AI assessment technologies are relatively limited (Chen et al. 2025), in that they are primarily confined to evaluation tasks. In contrast, as an emerging form of generative AI, ChatGPT offers a wider range of functionalities, such as text generation and interactive communication (Gallent-Torres et al. 2023). These features enable ChatGPT to support more diverse learning scenarios, which may, in turn, contribute to its larger effect size.

The meta-analysis results show that ChatGPT effectively enhances student learning performance across different types of courses, as has been shown by prior studies (Ba et al. 2024; Li, 2023; Song and Song, 2023). The most significant impact was observed for courses focused on skills and competencies development. This may be because such courses typically involve well-defined task objectives and procedural steps; ChatGPT

provides immediate feedback, targeted guidance, and problem-solving support (Rakap, 2023; Urban et al. 2024), and thus can enable students to complete such tasks more efficiently.

The present study reveals how students' learning patterns influence the positive impact of ChatGPT on their academic performance. Specifically, of the various learning models, ChatGPT demonstrates the most significant effect when applied to problem-based learning. This may be attributed to its inherent strengths, in that ChatGPT excels at providing clear explanations for a wide range of problems, whether establishing simple correspondences between geometric structures and algebraic equations (Poola and Božić, 2023) or explaining complex scientific concepts (Ali et al. 2024). ChatGPT thus allows students to access ample information that can enable them to solve teacher-assigned problems more efficiently, ultimately enhancing their learning performance. Moreover, this study yielded a unique conclusion: in project-based learning, the effect of ChatGPT on students' learning performance is the weakest, showing only a small positive impact. This may be because project-based learning emphasizes the completion of a comprehensive project in a real-world context (Botha, 2010). In such a scenario, ChatGPT's support for addressing complex problem-solving tasks may be limited, resulting in only a minor effect.

This meta-analysis revealed that ChatGPT has the most significant impact on learning performance when the duration of the learning experience is in the range of four to eight weeks. This underscores the importance of designing technological interventions to have an appropriate duration for optimal learning outcomes. This viewpoint is consistent with the conclusions of prior studies (Sung et al. 2017; Tian and Zheng, 2023). Interestingly, the current study found that the extent of ChatGPT's impact on student learning is minimized when the duration is less than one week. This may be due to the fact that when students first started using ChatGPT, they lacked the questioning skills needed to use it effectively. White et al. (2023) stated that the quality of ChatGPT output depends on the quality of the input. If students struggle to ask high-quality questions, then the quality of ChatGPT-generated text will suffer. This may be why ChatGPT has less effect on improving students' learning performance in the short term. Our results further indicate that when the usage period exceeds eight weeks, the positive effect of ChatGPT on students' learning performance slightly declines. This may be because the prolonged use of ChatGPT can lead students to become overly reliant on the AI tool, and thus they may neglect to reinforce the knowledge they have learned (Agarwal, 2023), which, in turn, would result in a decline in learning performance.

ChatGPT was found to have no significant difference in its impact on learning performance across different roles and application areas. This may be because ChatGPT provides such broad learning support, including extensive knowledge access, real-time interaction, personalized assistance, diverse resources, and data analysis services (Ali et al. 2024; Luo et al. 2024; Rawas, 2024; Yan et al. 2024). The various roles it can play and the areas in which it can be applied are essentially manifestations of its core function as a learning support tool. Therefore, these two moderating variables did not show significant differences with regard to learning performance.

How effective is ChatGPT in improving learning perception? Our results show that ChatGPT has a moderately positive effect on students' learning perceptions, with an overall effect size of 0.456. In addition, the impact of ChatGPT on students' learning perceptions is moderated by intervention duration. These meta-analytic findings show that ChatGPT is effective in fostering students' emotional attitudes toward learning, consistent with the results of prior studies (Chen and Chang, 2024; Lu et al. 2024; Urban et al. 2024). The technology acceptance model posits that perceived usefulness and perceived ease of use influence an individual's intention or attitude toward using new technology (Venkatesh and Bala, 2008). Therefore, the perceived usefulness and perceived ease of use of ChatGPT determine whether students will adopt it. ChatGPT can make a course more interesting and encourage students to actively search for information, ultimately contributing to an increase in learning perception. ChatGPT can meet student learning needs through actions such as providing instant feedback, offering personalized scaffolding of learning, and provoking interesting learning experiences (Li, 2023). Therefore, the students' learning perception is significantly improved. Notably, compared to its large positive impact on improving learning performance ($g = 0.867$), ChatGPT has only a moderately positive impact on learning perception. This may be because, although ChatGPT can quickly provide knowledge, answer questions, and generate content to enhance students' academic performance, it lacks emotional intelligence (Birenbaum, 2023), limiting its ability to provide the kind of "humanized" interaction needed to establish emotional resonance with students or spark deeper learning interest (Park and Ahn, 2024). Therefore, when guiding students to use ChatGPT as a learning aid, teachers should pay attention to their students' emotional and psychological needs to avoid overly mechanized and tool-driven learning experiences.

Of the six moderating variables examined, only intervention duration showed a significant effect ($p < 0.05$). More specifically, student learning perception increased with prolonged use of ChatGPT. This finding agrees with the results of Guan et al. (2024) and may be attributed to the consistent positive feedback students receive from ChatGPT during extended usage. Such feedback likely reinforces their positive emotions toward learning over time.

How effective is ChatGPT in fostering higher-order thinking? Our results indicate that ChatGPT has a moderately positive effect on the development of students' higher-order thinking, with an overall effect size of 0.457. Moreover, the effect of ChatGPT on higher-order thinking is moderated by the type of course and the role played by ChatGPT in the learning process. These meta-analytic findings suggest that ChatGPT can effectively foster students' higher-order thinking skills, in agreement with the results of prior studies (Li, 2023; Lu et al. 2024; Niloy et al. 2023). Lu et al. (2024) found that ChatGPT was able to help students reflect on their learning processes and strategies by providing immediate feedback and assessment, and could also help them formulate problem-solving ideas to enhance their metacognitive thinking and problem-solving skills. Woo et al. (2023) suggested that this could be attributed to ChatGPT's ability to present students with diverse information, thus promoting divergent and creative thinking. Metacognitive thinking, problem-solving abilities, and creative thinking are all crucial components of higher-order thinking. However, the improvements in learning performance attributed to ChatGPT are quite strong ($g = 0.867$), while the improvements in higher-order thinking are only moderate. This may be because ChatGPT can use simple language to express complex information (Niloy et al. 2023) and can provide students with a structured and logical framework for new knowledge (Li et al. 2024). This helps them master conceptual knowledge, and thus can rapidly improve students' learning performance. However, ChatGPT is based on pretrained models that rely on existing data and patterns. It thus may lack critical analysis abilities and the ability to provide

creative problem-solving solutions (Lu et al. 2024). Moreover, ChatGPT-generated content is not always completely accurate, and this bias can hinder the development of students' higher-order thinking (Niloy et al. 2023), thus limiting the overall degree to which ChatGPT can promote students' higher-order thinking.

The analysis of potential moderating variables shows that ChatGPT's effectiveness in fostering higher-order thinking differs significantly across different types of courses. Specifically, ChatGPT has the most substantial impact when used in STEM and related courses. This may be due to the goals of these courses. In STEM and related courses, students are often required to collaborate with ChatGPT to complete complex projects (Li, 2023; Li et al. 2024), and the process of completing such projects inherently promotes the development of higher-order thinking skills such as creative thinking, problem-solving, and critical thinking.

In terms of the roles that ChatGPT can play, its function as an intelligent tutor has a particularly strong impact on the development of students' higher-order thinking. When acting as an intelligent tutor, ChatGPT provides students with personalized guidance, feedback, and assessment (Escalante et al. 2023). When using ChatGPT, students continuously reflect on their learning processes and adjust their learning strategies in real-time. This process effectively fosters the development of students' higher-order thinking skills (Lu et al. 2024).

Conclusion and implications

Conclusion. This study used meta-analysis to analyze the impact of ChatGPT on student learning performance, learning perception, and higher-order thinking. With regard to learning performance, an analysis of data from 44 experimental and quasi-experimental studies showed that the calculated effect size indicates a large positive impact ($g = 0.867$) of ChatGPT on student learning performance. The analysis of moderating variables revealed no significant differences in the effect of ChatGPT across grade levels, role of ChatGPT, or area of ChatGPT application. However, significant differences were found in the type of course, learning model, and duration.

For learning perception, an analysis of data from 19 experimental and quasi-experimental studies revealed a medium positive effect ($g = 0.456$) of ChatGPT on student learning perception. The moderator analysis indicated that there were no significant differences in its impact across grade level, type of course, learning model, role of ChatGPT, or area of ChatGPT application. However, duration showed a significant effect.

In terms of higher-order thinking, data from nine experimental and quasi-experimental studies indicated a medium positive effect ($g = 0.457$) of ChatGPT on students' higher-order thinking. The analysis of moderating variables revealed no significant differences in the effect of ChatGPT across the learning model, duration, or area of the ChatGPT application. However, significant differences were observed in the type of course and role of ChatGPT.

In summary, the meta-analysis results of this study confirm the positive impacts of ChatGPT on learning performance, learning perception, and higher-order thinking, while highlighting some variations in effectiveness across different usage conditions and learning contexts. Based on these results, governments could introduce policies to promote the integration of ChatGPT into education, encouraging universities and secondary schools to apply ChatGPT appropriately. Additionally, ChatGPT could be actively incorporated into STEM and related courses, skills and competencies development training programs, and language learning and academic writing courses to enhance student learning.

Implications for theory and practice. This meta-analysis makes significant contributions to our understanding of the theoretical and practical aspects of the use of ChatGPT in education. From a theoretical perspective, it extends the existing literature on ChatGPT and student learning by providing quantitative evidence that ChatGPT can indeed effectively improve learning performance, enhance learning perception, and foster higher-order thinking. Second, this study provides a more thorough and integrated investigation of potential moderating variables. It explores the effects of ChatGPT on student learning across different types of courses, grade levels, learning models, intervention durations, instructional roles, and application types. It has a broader scope than the meta-analyses by Heung and Chiu (2025) and Guan et al. (2024). Compared to Deng et al. (2025), this study further explores learning methods, the role of ChatGPT in learning, and its application across various domains. Its findings reveal that different learning methods and the roles played by ChatGPT significantly affect students' learning performance and higher-order thinking, resulting in notable differences in outcomes. Broadly speaking, this study provides a theoretical foundation for the future development of national AI education policies and the application of ChatGPT by educators, while further enriching the research on the use of ChatGPT in the field of education.

From a practical perspective, the results of the analysis of potential moderating variables in this study show that ChatGPT should not be implemented arbitrarily. Instead, it should be used scientifically and reasonably according to the type of course, learning model, duration, etc. Future applications of ChatGPT should pay attention to the following points: (1) When applying ChatGPT to developing students' higher-order thinking, it is crucial to provide corresponding learning scaffolds or educational frameworks (e.g., Bloom's taxonomy) because ChatGPT lacks creativity and critical thinking. Doing so can improve its effectiveness in cultivating higher-order thinking. (2) Broad cross-grade use should be encouraged to support student learning. ChatGPT can serve as an auxiliary learning tool for both middle school and university students, enhancing both their academic performance and their learning perception. In middle schools, teachers can use ChatGPT to maintain students' interest while improving their understanding and memory of complex concepts. In universities, teachers can guide students to actively use ChatGPT in relevant courses to provide personalized learning suggestions to improve their learning. (3) ChatGPT can be actively applied in different types of courses to enhance student learning. For example, in STEM and related courses, ChatGPT can be used to present diverse information to help students think divergently and complete projects. Alternatively, ChatGPT can be used to simulate problem-based learning environments by generating scenario-based questions. In skills and competencies development training, ChatGPT can be used to provide more targeted guidance. In language learning and academic writing courses, ChatGPT can help students enhance their comprehension of texts and improve the accuracy and grammatical correctness of their writing. (4) ChatGPT can be applied in different learning models, particularly in problem-based learning, as its use in this mode best supports students' learning performance. (5) Continuous use of ChatGPT can be ensured to support student learning, with a recommended duration of four to eight weeks. For shorter durations, teachers should provide relevant scaffolding, such as guidance on how to write high-quality ChatGPT prompts. If a longer duration is necessary, teachers can consider introducing new instructional models, such as blended learning or flipped classrooms, and apply ChatGPT more strategically in different teaching stages such as pre-class preparation, in-class interaction, or post-class review to enhance

learning outcomes. Moreover, teachers can conduct regular formative assessments to monitor students' progress and identify gaps in their understanding. This would help prevent the students from developing an over-reliance on technology that could hinder their knowledge retention, while also promoting deeper learning and long-term mastery. (6) ChatGPT should be used flexibly in teaching as an intelligent tutor, partner, learning tool, and more. Its role(s) can be adjusted according to the requirements of the teaching and learning activities to which it is applied.

Limitations and future research

There are several limitations that should be solved for future work. First, this study sample size was limited, comprising only 51 documents, of which only one addressed primary school and none addressed kindergarten. Moreover, the number of included samples that examined learning perception and higher-order thinking was relatively small, with only 19 and 9 studies, respectively, compared to a significantly larger sample for learning performance. This imbalance may limit our ability to comprehensively understand ChatGPT's impact on learning perception and higher-order thinking. Therefore, future work could include more studies. Second, this study only explored seven moderating variables, but there are many more variables that could potentially affect the impact of ChatGPT on student learning, such as students' cultural backgrounds, parents' occupations, and relevant national policies. Future studies should address a broader scope of factors. Finally, this study only selected experimental and quasi-experimental studies and did not include relevant qualitative or mixed-methods research. Future review studies could integrate a broader array of related studies, including qualitative and mixed-methods research, to analyze the impact of ChatGPT on student learning from a more comprehensive and holistic perspective, thereby broadening the scope and applicability of the findings.

Data availability

No datasets were generated or analysed during the current study.

Received: 1 December 2024; Accepted: 19 March 2025;
Published online: 06 May 2025

References

*An S, Zhang S, Guo T, Lu S, Zhang W, Cai Z (2025) Impacts of generative AI on student teachers' task performance and collaborative knowledge construction process in mind mapping-based collaborative environment. Comput Educ 227. https://doi.org/10.1016/j.compedu.2024.105227

*Avello-Martínez R, Gajderowicz T, Gómez-Rodríguez VG (2024) Is ChatGPT helpful for graduate students in acquiring knowledge about digital storytelling and reducing their cognitive load? An experiment. Revista de Educación a Distancia 24(78):8. https://doi.org/10.6018/red.604621

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

*Bašic Ž, Banovac A, Kružic I, Jerkovic I (2023a) ChatGPT-3.5 as writing assistance in students' essays. Humanit Soc Sci Commun 10:750. https://doi.org/10.1057/s41599-023-02269-7

*Basic Z, Banovac A, Kruzic I, Jerkovic I (2023b) Better by you, better than me, chatgpt3 as writing assistance in students essays. arXiv. https://doi.org/10.48550/arXiv.2302.04536

*Boudouaia A, Mouas S, Kouider B (2024) A study on ChatGPT-4 as an Innovative approach to enhancing English as a foreign language writing learning. J Educ Comput Res. https://doi.org/10.1177/07356331241247465

*Chan S, Lo NPK, Wong A (2024) Generative AI and essay writing: impacts of automated feedback on revision performance and engagement. Reflections 31(3):1249–1284. https://doi.org/10.61508/refl.v31i3.277514

*Chen C-H, Chang C-L (2024) Effectiveness of AI-assisted game-based learning on science learning outcomes, intrinsic motivation, cognitive load, and learning behavior. Educ Inf Technol. https://doi.org/10.1007/s10639-024-12553-x

*Darmawansah D, Rachman D, Febiyani F, Hwang G-J (2024) ChatGPT-supported collaborative argumentation: integrating collaboration script and argument mapping to enhance EFL students' argumentation skills. Educ Inf Technol. https://doi.org/10.1007/s10639-024-12986-4

*Donald MJ, Will D, Christopher ME (2024) Using ChatGPT with novice arduino programmers: Effects on performance, interest, self-efficacy, and programming ability. J Res Tech Careers 8(1):1–17. https://doi.org/10.9741/2578-2118.1152

*Emran AQ, Mohammed MN, Saeed H, Abu Keir MY, Alani ZN, Mohammed Ibrahim F (2024) Paraphrasing ChatGPT answers as a tool to enhance university students' academic writing skills. In: 2024 ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems (ICETSIS). IEEE, pp. 501–505

*Escalante J, Pack A, Barrett A (2023) AI-generated feedback on writing: insights into efficacy and ENL student preference. Int J Educ Technol High Educ 20(1). https://doi.org/10.1186/s41239-023-00425-2

*Essel HB, Vlachopoulos D, Essuman AB, Amankwa JO (2024) ChatGPT effects on cognitive skills of undergraduate students: receiving instant responses from AI-based conversational large language models (LLMs). Comput Educ: Artif Intell 6. https://doi.org/10.1016/j.caeai.2023.100198

*Gan W, Ouyang J, Li H, Xue Z, Zhang Y, Dong Q, Huang J, Zheng X, Zhang Y (2024) Integrating ChatGPT in orthopedic education for medical undergraduates: randomized controlled trial. J Med Internet Res 26:e57037. https://doi.org/10.2196/57037

*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. Lang Teach Res. https://doi.org/10.1177/13621688241239764

*Hsu MH (2024) Mastering medical terminology with ChatGPT and Termbot. Health Educ J 83(4):352–358. https://doi.org/10.1177/00178969231197371

*Hu Y-H (2024) Implementing generative AI chatbots as a decision aid for enhanced values clarification exercises in online business ethics education. Educ Technol Soc 27(3):356–373. https://doi.org/10.30191/ETS.202407_27(3).TP02

*Huang K-L, Liu Y-c, Dong M-Q (2024) Incorporating AIGC into design ideation: A study on self-efficacy and learning experience acceptance under higher-order thinking. Thinking Skills and Creativity, 52. https://doi.org/10.1016/j.tsc.2024.101508

*Huesca G, Martínez-Treviño Y, Molina-Espinosa JM, Sanromán-Calleros AR, Martínez-Román R, Cendejas-Castro EA, Bustos R (2024) Effectiveness of using ChatGPT as a tool to strengthen benefits of the flipped learning strategy. Educ Sci 14(6):660. https://doi.org/10.3390/educsci14060660

*Jafarian NR, Kramer A-W (2025) AI-assisted audio-learning improves academic achievement through motivation and reading engagement. Comput Educ: Artif Intell 8. https://doi.org/10.1016/j.caeai.2024.100357

*Ji Y, Zou X, Li T, Zhan Z (2023) The effectiveness of ChatGPT on pre-service teachers' STEM teaching literacy, learning performance, and cognitive load in a teacher training course. *The 6th International Conference on Educational Technology Management*, Guangzhou, China

*Karaman MR, Goksu I (2024) Are lesson plans created by ChatGPT more effective? An experimental study. Int J Technol Edu 7(1):107–127. https://doi.org/10.46328/ijte.607

*Kareem MS, Elham GS, Sawsan HAK, Chai CT (2025) ChatGPT-Related Risk Patterns and Students' Creative Thinking Toward Tourism Statistics Course: Pretest and Posttest Quasi-Experimentation. J Hospitality Tourism Educ. https://doi.org/10.1080/10963758.2025.2456638

*Kosar T, Ostojic D, Liu YD, Mernik M (2024) Computer science education in ChatGPT era: Experiences from an experiment in a programming course for novice programmers. Mathematics 12(5):629. https://doi.org/10.3390/math12050629

*Küchemann S, Steinert S, Revenga N, Schweinberger M, Dinc Y, Avila KE, Kuhn J (2023) Can ChatGPT support prospective teachers in physics task development? Phys Rev Phys Educ Res 19(2):020128. https://doi.org/10.1103/PhysRevPhysEducRes.19.020128

*Li H-F (2023) Effects of a ChatGPT-based flipped learning guiding approach on learners' courseware project performances and perceptions. Australas J Educ Technol 39(5):40–58. https://doi.org/10.14742/ajet.8923

*Li T, Ji Y, Zhan Z (2024) Expert or machine? Comparing the effect of pairing student teacher with in-service teacher and ChatGPT on their critical thinking, learning performance, and cognitive load in an integrated-STEM course. Asia Pac J Educ 44(1):45–60. https://doi.org/10.1080/02188791.2024.2305163

*Liao J, Zhong L, Zhe L, Xu H, Liu M, Xie T (2024) Scaffolding computational thinking with ChatGPT. IEEE Trans Learn Technol 17:1668-1682. https://doi.org/10.1109/tlt.2024.3392896

*Lu J, Zheng R, Gong Z, Xu H (2024) Supporting teachers' professional development with generative AI: the effects on higher order thinking and self-

efficacy. IEEE Trans Learn Technol 17:1279–1289. https://doi.org/10.1109/tlt.2024.3369690
*Maghamil MC, Sieras SG (2024) Impact of ChatGPT on the academic writing quality of senior high school students. J Engl Lang Teach Appl Linguist 6(2):115–128. https://doi.org/10.32996/jeltal.2024.6.2.14
*Mahapatra S (2024) Impact of ChatGPT on ESL students' academic writing skills: A mixed methods intervention study. Smart Learn Environ 11(1):9. https://doi.org/10.1186/s40561-024-00295-9
*Mugableh AI (2024) The impact of ChatGPT on the development of vocabulary knowledge of Saudi EFL students. Arab World Engl J 265–281. https://doi.org/10.24093/awej/ChatGPT.18
*Niloy AC, Akter S, Sultana N, Sultana J, Rahman SIU (2023) Is ChatGPT a menace for creative writing ability? An experiment. J Comput Assist Learn 40(2):919–930. https://doi.org/10.1111/jcal.12929
*Pan M, Lai C, Guo K (2025) Effects of GenAI-empowered interactive support on university EFL students' self-regulated strategy use and engagement in reading. Internet High Educ 65. https://doi.org/10.1016/j.iheduc.2024.100991
*Qi L, Yuan Y, Longhai X, Mingzhu Y, Jue W, Xinhua Z (2024) Can ChatGPT effectively complement teacher assessment of undergraduate students' academic writing? Assess Eval High Educ 49(5):616–633. https://doi.org/10.1080/02602938.2024.2301722
*Roganovic J (2024) Familiarity with ChatGPT Features Modifies Expectations and Learning Outcomes of Dental Students. Int Dent J 74(6):1456–1462. https://doi.org/10.1016/j.identj.2024.04.012
*Shang S, Geng, S (2024) Empowering learners with AI-generated content for programming learning and computational thinking: The lens of extended effective use theory. J Comput Assist Learn 40:1941–1958. https://doi.org/10.1111/jcal.12996
*Sun D, Boudouaia A, Zhu C, Li Y (2024) Would ChatGPT-facilitated programming mode impact college students' programming behaviors, performances, and perceptions? An empirical study. Int J Educ Technol High Educ 21(1). https://doi.org/10.1186/s41239-024-00446-5
*Yang T-C, Hsu Y-C, Wu J-Y (2025) The effectiveness of ChatGPT in assisting high school students in programming learning: evidence from a quasi-experimental research. Interact Learn Environ. https://doi.org/10.1080/10494820.2025.2450659
*Urban M, Dechterenko F, Lukavský J, Hrabalová V, Svacha F, Brom C, Urban K (2024) ChatGPT improves creative problem-solving performance in university students: an experimental study. Comput Educ 215. https://doi.org/10.1016/j.compedu.2024.105031
*Wu T-T, Lee H-Y, Li P-H, Huang C-N, Huang Y-M (2024) Promoting self-regulation progress and knowledge construction in blended learning via ChatGPT-based learning aid. J Educ Comput Res 61(8):3–31. https://doi.org/10.1177/07356331231191125
*Yilmaz R, Karaoglan Yilmaz, FG (2023) The effect of generative artificial intelligence (AI)-based tool use on students' computational thinking skills, programming self-efficacy and motivation. Comput Educ: Artif Intell 4:100147. https://doi.org/10.1016/j.caeai.2023.100147
*Zhou W, Kim Y (2024) Innovative music education: an empirical assessment of ChatGPT-4's impact on student learning experiences. Educ Inf Technol 29, 20855–20881. https://doi.org/10.1007/s10639-024-12705-z
Abu Khurma O, Albahti F, Ali N, Bustanji A (2024) AI ChatGPT and student engagement: unraveling dimensions through PRISMA analysis for enhanced learning experiences. Contemp Educ Technol 16(2):ep503. https://doi.org/10.30935/cedtech/14334
Abuhassna H, Alnawajha S (2023b) The transactional distance theory and distance learning contexts: theory integration, research gaps, and future agenda. Educ Sci 13(2):112. https://doi.org/10.3390/educsci13020112
Abuhassna H, Alnawajha S (2023a) Instructional design made easy! Instructional design models, categories, frameworks, educational context, and recommendations for future work. Eur J Investig Health Psychol Educ 13(4):715–735. https://doi.org/10.3390/ejihpe13040054
Abuhassna H, Yahaya N, Megat zakaria MAZ, abu samah N, Zaid N, Awae F, Chee KN, Alsharif A (2023) Trends on using the technology acceptance model (TAM) for online learning: a bibliometric and content analysis. Int J Inf Educ Technol 13:131–142. https://doi.org/10.18178/ijiet.2023.13.1.1788
Agarwal, K (2023). AI in education—evaluating ChatGPT as a virtual teaching assistant. Int J Multidiscipl Res 5(4). https://doi.org/10.36948/ijfmr.2023.v05i04.4484
Albdrani RN, Al-Shargabi AA (2023) Investigating the effectiveness of ChatGPT for providing personalized learning experience: a case study. Int J Adv Comput Sci Appl 14(11):1208–1213. https://doi.org/10.14569/IJACSA.2023.01411122
Ali D, Fatemi Y, Boskabadi E, Nikfar M, Ugwuoke J, Ali H (2024) ChatGPT in teaching and learning: a systematic review. Educ Sci 14(6):643. https://doi.org/10.3390/educsci14060643
Almohesh ARI (2024) AI application (ChatGPT) and Saudi Arabian primary school students' autonomy in online classes: exploring students and teachers' perceptions. Int Rev Res Open Distrib Learn 25(3):7641. https://doi.org/10.19173/irrodl.v25i3.7641
Alneyadi S, Wardat Y (2023) ChatGPT: revolutionizing student achievement in the electronic magnetism unit for eleventh-grade students in Emirates schools. Contemp Educ Technol 15(4). https://doi.org/10.30935/cedtech/13417
Alsharif MHK, Elamin AY, Almasaad JM, Bakhit NM, Alarifi A, Taha KM, Hassan WA, Zumrawi E (2024) Using ChatGPT to create engaging problem-based learning scenarios in anatomy: a step-by-step guide. New Armen Med J 18(4). https://doi.org/10.56936/18290825-4.v18.2024-98
Ansari AN, Ahmad S, Bhutta SM (2024) Mapping the global evidence around the use of ChatGPT in higher education: a systematic scoping review. Educ Inf Technol 29(9):11281–11321. https://doi.org/10.1007/s10639-023-12223-4
Avsheniuk N, Lutsenko O, Svyrydiuk T, Seminikhyna N (2024) Empowering language learners' critical thinking: evaluating ChatGPT's role in English course implementation. Arab World Engl J 210–224. https://doi.org/10.24093/awej/ChatGPT.14
Baig MI, Yadegaridehkordi E (2024) ChatGPT in the higher education: a systematic literature review and research challenges. Int J Educ Res 127:102411. https://doi.org/10.1016/j.ijer.2024.102411
Beltozar-Clemente S, Díaz-Vega E (2024) Physics XP: integration of ChatGPT and gamification to improve academic performance and motivation in Physics 1 course. Int J Eng Pedagog 14(6):82–92. https://doi.org/10.3991/ijep.v14i6.47127
Birenbaum M (2023) The Chatbots' challenge to education: disruption or destruction? Educ Sci 13(7):711. https://doi.org/10.3390/educsci13070711
Bitzenbauer P (2023) ChatGPT in physics education: a pilot study on easy-to-implement activities. Contemp Educ Technol 15(3):ep430. https://doi.org/10.30935/cedtech/13176
Borenstein M, Hedges LV, Higgins JPT, Rothstein HR (2009) Introduction to meta-analysis. John Wiley & Sons
Botha M (2010) A project-based learning approach as a method of teaching entrepreneurship to a large group of undergraduate students in South Africa. Educ Change 14(2):213–232. https://doi.org/10.1080/16823206.2010.522059
Chen A, Zhang Y, Jia J, Liang M, Cha Y, Lim CP (2025) A systematic review and meta-analysis of AI-enabled assessment in language learning: design, implementation, and effectiveness. J Comput Assist Learn 41(1):e13064. https://doi.org/10.1111/jcal.13064
Chen X, Hu Z, Wang C (2024) Empowering education development through AIGC: a systematic literature review. Educ Inf Technol 29(13):17485–17537. https://doi.org/10.1007/s10639-024-12549-7
Coban A, Dzsotjan D, Küchemann S, Durst J, Kuhn J, Hoyer C (2025) AI support meets AR visualization for Alice and Bob: personalized learning based on individual ChatGPT feedback in an AR quantum cryptography experiment for physics lab courses. EPJ Quantum Technol 12(1):15. https://doi.org/10.1140/epjqt/s40507-025-00310-z
Dempere J, Modugu K, Hesham A, Ramasamy LK (2023) The impact of ChatGPT on higher education. Front Educ 8:1206936. https://doi.org/10.3389/feduc.2023.1206936
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. Comput Educ 227:105224. https://doi.org/10.1016/j.compedu.2024.105224
Ding L, Li T, Jiang SY, Gapud A (2023) Students' perceptions of using ChatGPT in a physics class as a virtual tutor. Int J Educ Technol High Educ 20(1):63. https://doi.org/10.1186/s41239-023-00434-1
Divito CB, Katchikian BM, Gruenwald JE, Burgoon JM (2024) The tools of the future are the challenges of today: the use of ChatGPT in problem-based learning medical education. Med Teach 46(3):320–322. https://doi.org/10.1080/0142159x.2023.2290997
Economides AA, Perifanou M (2024) University students using ChatGPT in project-based learning. In: Maria Perifanou & Anastasios A. Economides (eds.), Digital transformation in higher education. empowering teachers and students for tomorrow's challenges. Springer Nature, pp. 27–39
Farah JC, Ingram S, Spaenlehauer B, Lasne FK-L, Gillet D (2023) Prompting large language models to power educational Chatbots. In: Xie H, Lai, CL., Chen W, Xu G, Popescu, E. (eds.), Advances in web-based learning—ICWL 2023. Springer Nature, pp. 169–188
Fokides E, Peristeraki E (2025) Comparing ChatGPT's correction and feedback comments with that of educators in the context of primary students' short essays written in English and Greek. Educ Inf Technol 30(2):2577–2621. https://doi.org/10.1007/s10639-024-12912-8
Gallent-Torres C, Zapata-González A, Ortego-Hernando JL (2023) The impact of Generative Artificial Intelligence in higher education: a focus on ethics and academic integrity. Relieve-Rev Electron Investig Eval Educ 29(2). https://doi.org/10.30827/relieve.v29i2.29134
Gouia-Zarrad R, Gunn C (2024) Enhancing students' learning experience in mathematics class through ChatGPT. Int Electron J Math Educ 19(3):em0781. https://doi.org/10.29333/iejme/14614

Grassini S (2023) Shaping the future of education: exploring the potential and consequences of AI and ChatGPT in educational settings. Educ Sci 13(7):692. https://doi.org/10.3390/educsci13070692
Guan L, Li S, Gu MM (2024) AI in informal digital English learning: a meta-analysis of its effectiveness on proficiency, motivation, and self-regulation. Comput Educ: Artif Intell 7:100323. https://doi.org/10.1016/j.caeai.2024.100323
Haindl P, Weinberger G (2024) Students' experiences of using ChatGPT in an undergraduate programming course. IEEE Access 12:43519–43529. https://doi.org/10.1109/access.2024.3380909
Hamid H, Zulkifli K, Naimat F, Yaacob NLC, Ng KW (2023) Exploratory study on student perception on the use of chat AI in process-driven problem-based learning. Curr Pharm Teach Learn 15(12):1017–1025. https://doi.org/10.1016/j.cptl.2023.10.001
Hartley K, Hayak M, Ko UH (2024) Artificial intelligence supporting independent student learning: an evaluative case study of ChatGPT and learning to code. Educ Sci 14(2):120. https://doi.org/10.3390/educsci14020120
Hays L, Jurkowski O, Sims SK (2024) ChatGPT in K-12 education. Techtrends 68(2):281–294. https://doi.org/10.1007/s11528-023-00924-z
Hedges LV (1981) Distribution theory for glass's estimator of effect size and related estimators. J Educ Stat 6(2):107. https://doi.org/10.2307/1164588
Heung YME, Chiu TKF (2025) How ChatGPT impacts student engagement from a systematic review and meta-analysis study. Comput Educ: Artif Intell 8. https://doi.org/10.1016/j.caeai.2025.100361
Higgins JP, Thompson SG, Deeks JJ, Altman DG (2003) Measuring inconsistency in meta-analyses. BMJ 327(7414):557–560. https://doi.org/10.1136/bmj.327.7414.557
Hui Z, Zewu Z, Jiao H, Yu C (2025) Application of ChatGPT-assisted problem-based learning teaching method in clinical medical education. BMC Med Educ 25(1):50. https://doi.org/10.1186/s12909-024-06321-1
Imran M, Almusharraf N (2023) Analyzing the role of ChatGPT as a writing assistant at higher education level: a systematic review of the literature. Contemp Educ Technol 15:1–14. https://doi.org/10.30935/cedtech/13605
Jalil S, Rafi S, LaToza TD, Moran K, Lam W (2023) ChatGPT and software testing education: promises & perils. In: 2023 IEEE International Conference on Software Testing, Verification and Validation Workshops (ICSTW). IEEE, pp. 4130–4137
Jauhiainen JS, Guerra AG (2024) Generative AI and education: dynamic personalization of pupils' school learning material with ChatGPT. Front Educ 9:1288723. https://doi.org/10.3389/feduc.2024.1288723
Jošt G, Taneski V, Karakatic S (2024) The impact of large language models on programming education and student learning outcomes. Appl Sci 14(10):4115. https://doi.org/10.3390/app14104115
Kasneci E, Sessler K, Küchemann S, Bannert M, Dementieva D, Fischer F, Gasser U, Groh G, Günnemann S, Hüllermeier E, Krusche S, Kutyniok G, Michaeli T, Nerdel C, Pfeffer J, Poquet O, Sailer M, Schmidt A, Seidel T, Kasneci G (2023) ChatGPT for good? On opportunities and challenges of large language models for education. Learn Individ Differ 103:102274. https://doi.org/10.1016/j.lindif.2023.102274
Kotsis KT (2024a) ChatGPT in teaching physics hands-on experiments in primary school. Eur J Educ Stud 11(10). https://doi.org/10.46827/ejes.v11i10.5549
Kotsis KT (2024b) Integrating ChatGPT into the Inquiry-based science curriculum for primary education. Eur J Educ Pedagog 5(6). https://doi.org/10.24018/ejedu.2024.5.6.891
Leite BS (2023) Artificial intelligence and chemistry teaching: a propaedeutic analysis of ChatGPT in chemical concepts defining br. Quim Nov 46(9):915–923. https://doi.org/10.21577/0100-4042.20230059
Levine S, Beck SW, Mah C, Phalen L, PIttman J (2025) How do students use ChatGPT as a writing support? J Adolesc Adult Lit 68(5):445–457. https://doi.org/10.1002/jaal.1373
Lin X (2024) Exploring the role of ChatGPT as a facilitator for motivating self-directed learning among adult learners. Adult Learn 35(3):156–166. https://doi.org/10.1177/10451595231184928
Lipsey MW, Wilson DB (2001) Practical meta-analysis. Sage Publications, Inc
Lo CK, Hew KF, Jong MS-Y (2024) The influence of ChatGPT on student engagement: a systematic review and future research agenda. Comput Educ 219. https://doi.org/10.1016/j.compedu.2024.105100
Looi CK, Jia FL (2025) Personalization capabilities of current technology chatbots in a learning environment: an analysis of student–tutor bot interactions. Educ Inf Technol. https://doi.org/10.1007/s10639-025-13369-z
Luo H, Liao X, Ru Q, Wang Z (2024) Generative AI-supported teacher comments: an empirical study based on junior high school mathematics classrooms. e-Educ Res 45(5):58–66. https://doi.org/10.13811/j.cnki.eer.2024.05.008
Mai DTT, Da CV, Hanh NV (2024) The use of ChatGPT in teaching and learning: a systematic review through SWOT analysis approach. Front Educ 9. https://doi.org/10.3389/feduc.2024.1328769
Masrom MB, Busalim AH, Abuhassna H, Mahmood NHN (2021) Understanding students' behavior in online social networks: a systematic literature review. Int J Educ Technol High Educ 18(1):6. https://doi.org/10.1186/s41239-021-00240-7
Maurya RK (2024) A qualitative content analysis of ChatGPT's client simulation role-play for practising counselling skills. Couns Psychother Res 24(2):614–630. https://doi.org/10.1002/capr.12699
Moher D, Liberati A, Tetzlaff J, Altman DG (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 151(4):264–269. https://doi.org/10.7326/0003-4819-151-4-200908180-00135
Moneus A, Al-Wasy BQ (2024) The impact of artificial intelligence on the quality of Saudi translators' performance. Al-Andal J Humanit Soc Sci 11(96):201–230. https://doi.org/10.35781/1637-000-096-006
Ng DTK, Tan CW, Leung JKL (2024) Empowering student self-regulated learning and science education through ChatGPT: a pioneering pilot study. Br J Educ Technol 55(4):1328–1353. https://doi.org/10.1111/bjet.13454
Nugroho A, Putro N, Syamsi K (2023) The potentials of ChatGPT for language learning: unpacking its benefits and limitations. Register J 16(2):224–247. https://doi.org/10.18326/rgt.v16i2.224-247
Nugroho A, Andriyanti E, Widodo P, Mutiaraningrum I (2024) Students' appraisals post-ChatGPT use: students' narrative after using ChatGPT for writing. Innov Educ Teach Int 1–13. https://doi.org/10.1080/14703297.2024.2319184
OpenAI (2022) Introducing ChatGPT. OpenAI
Page MJ, Moher D (2017) Evaluations of the uptake and impact of the preferred reporting items for Systematic reviews and Meta-Analyses (PRISMA) statement and extensions: a scoping review. Syst Rev 6(1):1–14. https://doi.org/10.1186/s13643-017-0663-8
Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, Chou R, Glanville J, Grimshaw JM, Hrobjartsson A, Lalu MM, Li T, Loder EW, Mayo-Wilson E, McDonald S, Moher D (2021) The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 372:n71. https://doi.org/10.1136/bmj.n71
Park H, Ahn D (2024) The promise and peril of ChatGPT in higher education: opportunities, challenges, and design implications. In: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, Honolulu, HI, USA
Pavlenko O, Syzenko A (2024) Using ChatGPT as a learning tool: a study of Ukrainian students' perceptions. Arab World English J 252–264. https://doi.org/10.24093/awej/ChatGPT.17
Piaget J (1970) Structuralism. Basic Books
Polakova P, Ivenz P (2024) The impact of ChatGPT feedback on the development of EFL students' writing skills. Cogent Educ 11(1):2410101. https://doi.org/10.1080/2331186x.2024.2410101
Poola I, Božic V (2023) Guiding AI with human intuition for solving mathematical problems in ChatGPT. Int J Eng Sci Res 11(07):31–40. https://www.researchgate.net/publication/373447147
Rahimi AR, Sheyhkholeslami M, Pour AM (2025) Uncovering personalized L2 motivation and self-regulation in ChatGPT-assisted language learning: a hybrid PLS-SEM-ANN approach. Comput Hum Behav Rep 17:100539. https://doi.org/10.1016/j.chbr.2024.100539
Rahman MM, Watanobe Y (2023) ChatGPT for education and research: opportunities, threats, and strategies. Appl Sci-Basel 13(9):5783. https://doi.org/10.3390/app13095783
Rakap S (2023) Chatting with GPT: enhancing individualized education program goal development for novice special education teachers. J Spec Educ Technol 39(3):339–348. https://doi.org/10.1177/01626434231211295
Rapti S, Sapounidis T, Tselegkaridis S (2023) Enriching a traditional learning activity in preschool through augmented reality: children's and teachers' views. Information 14(10):530. https://doi.org/10.3390/info14100530
Rawas S (2024) ChatGPT: empowering lifelong learning in the digital age of higher education. Educ Inf Technol 29(6):6895–6908. https://doi.org/10.1007/s10639-023-12114-8
Saleem N, Mufti T, Sohail SS, Madsen DO (2024) ChatGPT as an innovative heutagogical tool in medical education. Cogent Educ 11(1):2332850. https://doi.org/10.1080/2331186x.2024.2332850
Samsul SA, Yahaya N, Abuhassna H (2023) Education big data and learning analytics: a bibliometric analysis. Humanit Soc Sci Commun 10(1):709. https://doi.org/10.1057/s41599-023-02176-x
Sawilowsky SS (2009) New effect size rules of thumb. J Mod Appl Stat Methods. https://doi.org/10.22237/jmasm/1257035100
Schulz KF, Chalmers I, Hayes RJ, Altman DG (1995) Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA 273(5):408–412. https://doi.org/10.1001/jama.273.5.408
Seth I, Lim B, Cevik J, Sofiadellis F, Ross RJ, Cuomo R, Rozen WM (2024) Utilizing GPT-4 and generative artificial intelligence platforms for surgical education: an experimental study on skin ulcers. Eur J Plast Surg 47(1):19. https://doi.org/10.1007/s00238-024-02162-9
Song C, Song Y(2023) Enhancing academic writing skills and motivation: assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students [Article] Front Psychol 14:1260843. https://doi.org/10.3389/fpsyg.2023.1260843

Su Y, Lin Y, Lai C (2023) Collaborating with ChatGPT in argumentative writing classrooms. Assess Writ 57:100752. https://doi.org/10.1016/j.asw.2023.100752
Sung YT, Yang JM, Lee HY (2017) The effects of mobile-computer-supported collaborative learning: meta-analysis and critical synthesis. Rev Educ Res 87(4):768–805. https://doi.org/10.3102/0034654317704307
Tian Q, Zheng X (2023) Effectiveness of online collaborative problem-solving method on students' learning performance: a meta-analysis. J Comput Assist Learn 40(1):326–341. https://doi.org/10.1111/jcal.12884
Tsai CY, Lin YT, Brown IK (2024) Impacts of ChatGPT-assisted writing for EFL English majors: feasibility and challenges. Educ Inf Technol 29(17): 22427–22445. https://doi.org/10.1007/s10639-024-12722-y
Venkatesh V, Bala H (2008) Technology acceptance model 3 and a research agenda on interventions. Decis Sci 39(2):273–315. https://doi.org/10.1111/j.1540-5915.2008.00192.x
Villan F, Santos RPd (2023) ChatGPT as Co-Advisor in Scientific Initiation: Action Research with Project-Based Learning in Elementary Education. Acta Scientiae, 25, 60–117. https://doi.org/10.17648/acta.scientiae.7474
Vygotskij LS, Hanfmann E (1966) Thought and language. MIT Press
Wang WS, Lin CJ, Lee HY, Huang YM, Wu TT (2024a) Integrating feedback mechanisms and ChatGPT for VR-based experiential learning: impacts on reflective thinking and AIoT physical hands-on tasks. Interact Learn Environ. https://doi.org/10.1080/10494820.2024.2375644
Wang Y, Liu W, Yu X, Li B, Wang Q (2024b) The impact of virtual technology on students' creativity: a meta-analysis. Comput Educ 215. https://doi.org/10.1016/j.compedu.2024.105044
White J, Fu Q, Hays S, Sandborn M, Olea C, Gilbert H, Elnashar A, Spencer-Smith J, Schmidt DC (2023) A prompt pattern catalog to enhance prompt engineering with ChatGPT. In: Proceedings of the 30th Conference on Pattern Languages of Programs. The Hillside Group, Monticello, IL, USA
Woo DJ, Wang Y, Susanto H, Guo K (2023) Understanding English as a foreign language students' idea generation strategies for creative writing with natural language generation tools. J Educ Comput Res 1–19. https://doi.org/10.1177/07356331231175999
Xiao YY, Zhi YY (2023) An exploratory study of EFL learners' use of ChatGPT for language learning tasks: experience and perceptions. Languages 8(3):212. https://doi.org/10.3390/languages8030212
Yan L, Sha L, Zhao L, Li Y, Martinez-Maldonado R, Chen G, Li X, Jin Y, Gaševic D (2024) Practical and ethical challenges of large language models in education: a systematic scoping review. Br J Educ Technol 55(1):90–112. https://doi.org/10.1111/bjet.13370
Young JC, Shishido M (2023) Investigating OpenAI's ChatGPT potentials in generating Chatbot's dialogue for English as a foreign language learning. Int J Adv Comput Sci Appl 14(6):65–72. https://doi.org/10.14569/IJACSA.2023.0140607
Yu CC, Yan JZ, Cai N (2024) ChatGPT in higher education: factors influencing ChatGPT user satisfaction and continued use intention. Front Educ 9:1354929. https://doi.org/10.3389/feduc.2024.1354929
Zhang X, Chen Y, Hu L, Wang Y (2022) The metaverse in education: definition, framework, features, potential applications, challenges, and future research topics. Front Psychol 13:1016300. https://doi.org/10.3389/fpsyg.2022.1016300
Zhang P, Tur G (2024) A systematic review of ChatGPT use in K-12 education. Eur J Educ 59(2). https://doi.org/10.1111/ejed.12599
Zhang Y, Liang Y, Tian X, Yu X (2024) The effects of unplugged programming activities on K-9 students' computational thinking: meta-analysis. Educ Technol Res Dev. https://doi.org/10.1007/s11423-023-10339-5
Zheng L, Niu J, Zhong L, Gyasi JF (2021) The effectiveness of artificial intelligence on learning achievement and learning perception: a meta-analysis. Interact Learn Environ 31(9):5650–5664. https://doi.org/10.1080/10494820.2021.2015693

Acknowledgement

This work was supported by the Major Project of Humanities and Social Sciences Key Research Base of the Ministry of Education (Grant No. 22JJD880023), University-Industry Collaborative Education Program (Grant No. 240904171111322), Hangzhou Social Science Routine Planning Project (Grant No. Z24JC014) and Hangzhou Normal University's 2023 Teaching Construction and Reform Project "Exploration and Practice of Generative Artificial Intelligence to Promote University Teaching Reform".

Author contributions

JW designed the research plan, was responsible for coding studies, experimental design, and data collection, participated in data analysis and interpretation, as well as writing and revising the initial manuscript. WF proposed the research topic, played an important guiding role throughout the research project, and made significant contributions to manuscript writing, revision, review, and polishing.

Competing interests

The authors declare no competing interests.

Ethical approval

This article does not include any studies performed by any of the authors with human participants or animals.

Informed consent

This article does not include any studies performed by any of the authors with human participants or animals.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1057/s41599-025-04787-y.

Correspondence and requests for materials should be addressed to Wenxiang Fan.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2025