Review

Journal of Educational Computing Research
2026, Vol. 64(3) 692–723

DOI: 10.1177/07356331251400540
journals.sagepub.com/home/jec
Sage

Do Generative AI-Powered Pedagogical Agents Improve Learners' Academic Performance Effectively? Evidence From Meta-Analysis

Liang Cheng[1], Hui Shi[2], Yuhan Wu[1], and Feng Li[1]

Abstract

With the breakthrough advancements in generative artificial intelligence (GenAI) technology, GenAI-powered pedagogical agents (GenAI-PA) are emerging as a transformative paradigm shift in education. However, there is still debate about whether the use of GenAI-PA is beneficial for learners' academic performance. Therefore, this study conducted a meta-analysis of 27 experimental and quasi-experimental studies from 2015 to 2025. The results showed that (1) GenAI-PA had a significant effect on learners' academic performance (g = 0.401). (2) In a collectivist culture, GenAI-PA had a greater effect on learners' academic performance. (3) The effect of GenAI-PA in teacher-directed learning was significantly better than in self-directed learning. (4) The dialogue modality of GenAI-PA moderated the effect on learners' academic performance, with multimodal dialogue showing the highest pedagogical potential. (5) The predictive effect of GenAI-PA on learners' academic performance was not influenced by grade level, gender, learning domain, learning duration, or the instructional role of GenAI-PA. Ultimately, recommendations for the design and application of GenAI-PA are discussed.

[1]Department of Education Information Technology, Faculty of Education, East China Normal University, Shanghai, China
[2]Faculty of Artificial Intelligence in Education, Central China Normal University, Wuhan, China

Corresponding Author:
Feng Li, Department of Education Information Technology, Faculty of Education, East China Normal University, Shanghai North Zhongshan Road Campus: 3663 N. Zhongshan Rd, Shanghai 200062, China.
Email: lifengys2001@163.com

Keywords
human-computer interaction, generative artificial intelligence, pedagogical agents, academic performance, meta-analysis

Introduction

The breakthrough development of generative artificial intelligence technology has positioned Generative AI-powered pedagogical agents (GenAI-PA) as a central focus in intelligent education research, serving as a key vehicle for technology-enhanced learning (Wei et al., 2024). GenAI-PA plays a virtual instructional role constructed upon Large Multimodal Models (LMMs) and intelligence technologies, providing learners with personalized cognitive support and emotional interaction through multiple channels (Bozkurt, 2023; Joshi, 2025; Zhang et al., 2025). The core objective of GenAI-PA is to simulate roles like teachers and learning partners, aiming to optimize learners' cognitive processing, stimulate learning motivation, and boost academic performance (Arguedas et al., 2024). In recent years, GenAI-PA has been extensively applied in scenarios such as project-based learning, online tutoring, and experimental teaching (Dai et al., 2025; Kumar, 2021; Ruiz Viruel et al., 2025). This highlights the potential of AI in optimizing learning experiences and enhancing instructional interaction (Kuhail et al., 2023).

Previous empirical studies indicated that GenAI-PA can improve learners' comprehension ability by in-depth situational dialogue interaction, and reduce learning anxiety through incentive scaffolding, consequently leading to better academic performance (e.g., Cheon et al., 2025; Lin & Yu, 2025). However, the existing researches have yet to reach a consensus on whether GenAI-PA can effectively enhance learners' academic performance. Some studies have revealed issues such as insufficient decision-making accuracy, superficial emotional interaction, and lack of structured guidance in GenAI-PA may undermine the advantages of intelligent technology (Alneyadi & Wardat, 2023; Pan et al., 2025). GenAI-PA risks being reduced to a mere answering tool, shifting learners from active knowledge constructors to passive recipients of information. This shift could undermine learners' abilities in independent thinking, critical analysis, and creative problem-solving (Mohammed et al., 2025). Consequently, GenAI-PA may hurt academic performance, particularly among learners who lack sufficient metacognitive skills to critically engage with AI-generated content (Yan et al., 2024).

Despite growing interest in applying GenAI-PA to enhance student outcomes, its effectiveness in improving academic performance remains uncertain. Additionally, the moderating effects of learner characteristics, instructional characteristics, and GenAI-PA characteristics remain undetermined (Bozkurt, 2023). This meta-analysis aims to systematically analyze relevant experimental and quasi-experimental studies from the last ten years, quantitatively synthesize the effect size of GenAI-PA in learning, and investigate influential factors through moderator analysis. Specifically, this study mainly focuses on the following two research questions:

RQ1: Can GenAI-PA effectively promote learners' academic performance?

RQ2: What variables mediate the promoting effect of GenAI-PA on learners' academic performance?

Literature Review

Generative AI-Powered Pedagogical Agents

The concept of pedagogical agents (PA) was proposed in the 1990s, referring to virtual instructional roles embedded in a digital learning environment, aiming to promote learners' cognitive learning (Makransky et al., 2019; Zhang et al., 2024). The surge of PA-related research in educational technology arose in response to a fundamental contradiction in education: the imbalance in the teacher-student ratio resulted in a scarcity of individualized guidance (Bloom, 1984). Consequently, it became challenging for human teachers to offer continuous metacognitive training and emotional support to every student (Efklides, 2006). With the advantages of replicability, scalability, and stability, PA assisted teachers in dealing with repetitive tasks and promoted the balance between large-scale education and individual needs (Apoki et al., 2022). In the initial period, PA was mostly used for structured tasks such as math problem-solving and language training, which were limited by static knowledge bases and fixed interaction logic in rule engines. A typical intelligent system was the Reader-specific Expert Adaptation Program (REAP), which was able to provide students with language learning paths and feedback based on preset rules and student learning (Heilman et al., 2006). With the evolution of artificial intelligence technology, PA gradually evolves into intelligence. After 2010, Intelligent Tutoring Systems (ITS) based on machine learning emerged, such as the Cognitive Tutor (Matsuda et al., 2015). ITS can adjust task difficulty dynamically based on student error patterns. However, such systems still rely on manually annotated knowledge graphs, suffering from deficiencies like insufficient dialogue flexibility, text-based interaction, and weak emotional support (Mousavinasab et al., 2021; Schroeder et al., 2024).

As the product of a technology paradigm shift, the core breakthrough of GenAI-PA was the ability to generate natural dialogue and teaching content in real time according to the context, achieving the leap from "preset response" to "dynamic creation" (Mehdian & Turi, 2025). For example, GPT 4.0-based "Khanmigo" could simulate Socratic dialogue and guide students to solve math problems (Shetye, 2024). The multimodal agent "Socratica" could combine text, images, and avatars to explain abstract scientific concepts through contextualized presentations (Borges & Araújo, 2024). Compared with PA, GenAI-PA had three advantages: (1) The ability to understand context and create new content; (2) Support open domain discussion and emotional empathy, which alleviating learning anxiety; (3) Real-time feedback based on the thought chain, promote the development of higher order thinking ability by layer guidance (Bektik et al., 2024; Chugh et al., 2025). In other words, GenAI-PA could better simulate various roles in the real learning environment, achieving more real and effective learning interaction.

The Functioning Mechanism of GenAI-PA in Learning

The impact of GenAI-PA on student learning can be explained through a theoretical lens. Mayer et al. (2003) proposed the Social Agency Theory, indicating that features such as visual, text, images, and sounds in pedagogical agents can serve as social cues, triggering social responses from learners and leading them to view the pedagogical agents as social partners and engage in social interaction with them. When the social cues provided by GenAI-PA are matched to real human interactions, learners will activate the social conversational schema, simulating real conversations through the "question-response-feedback" cycle. Meanwhile, Schneider et al. (2022) proposed the cognitive-affective-social theory of learning in digital environments (CASTLE), further clarifying that a digital pedagogical agent effectively integrating social elements can stimulate learners' participation at the cognitive and emotional levels by providing emotion regulation support and timely feedback. Domagk (2010) constructed the Pedagogical Agents Conditions of Use model (PACU), and pointed out that the effect of PA is jointly influenced by internal factors and external factors. Internal factors included the functional and morphological design of PA, while external factors included the instructional application scenarios and the characteristics of learners.

Based on these theoretical perspectives, this study constructed the functioning mechanism of GenAI-PA in learning (see Figure 1). The functioning mechanism of GenAI-PA in learning can be summarized as follows: GenAI-PA provides reasonable social cues for learners in specific instructional scenarios and promotes dialogue interaction between learners and GenAI-PA. The quality and effect of interaction are influenced by GenAI-PA characteristics, learner characteristics, and instructional characteristics. When GenAI-PA is reasonably designed, it can provide learners with positive emotional regulation, activating social conversation schemas by providing social cues, thereby promoting deep cognitive processing. On the contrary, when

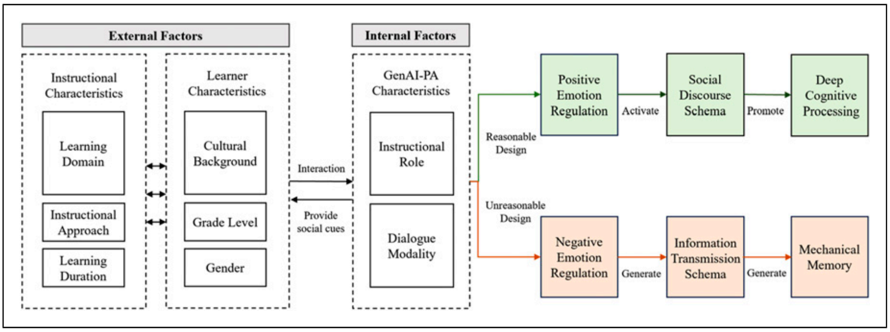

Figure 1. The Functioning Mechanism of GenAI-PA in Learning

GenAI-PA is unreasonably designed, negative emotion regulation will make learners believe that learning is merely a single information transmission. What learners experience is an information transmission schema, which leads to generate mechanical memory. For the rational design and application of GenAI-PA, educators need to have an in-depth understanding of the variables in the three aspects of GenAI-PA characteristics, learner characteristics, and instructional characteristics, as well as how these variables affect the cognitive development of learners.

Potential Moderating Variables

Previous research indicated that the effectiveness of GenAI-PA in promoting learning was influenced by multiple moderating factors (Bozkurt, 2023; Winkler et al., 2020). Based on the functioning mechanism of GenAI-PA we proposed, this study systematically investigated these moderators across these three dimensions.

Learner Characteristics

Cultural Background. Cultural values may shape students' engagement patterns and perceived pedagogical functions of GenAI-PA, thereby influencing their academic performance. According to Hofstede's cultural dimensions theory (Żemojtel-Piotrowska & Piotrowski, 2023), the collectivist culture tends to emphasize group harmony and teacher authority, leading students to regard GenAI-PA as supplementary tools while prioritizing interpersonal interaction with teachers. In contrast, the individualist culture encourages independent learning, which may foster greater reliance on the personalized guidance offered by GenAI-PA. These culturally influenced engagement patterns are reflected in empirical findings. Hooshyar et al. (2015) reported a significantly positive effect of GenAI-PA on programming achievement among Malaysian students with collectivist culture ($g = 0.858$), while Wang et al. (2025) observed a negative effect of GenAI-PA on programming performance among American students with individualist culture ($g = -0.415$). Therefore, we speculate that the correlation between GenAI-PA and academic performance is different in terms of cultural backgrounds.

Grade Level. Significant differences in learners' cognitive development across grade levels may influence their initiative and effectiveness in utilizing GenAI-PA. Xu et al. (2021) found that in reading learning activities of lower grades, students' initiative in using GenAI-PA was insufficient, resulting in a weaker teaching effect based on GenAI-PA than that guided by human teachers ($g = -0.024$). On the other hand, Tai et al. (2024) used Google Assistant to teach English listening and speaking to grade 9 students. The experimental group actively interacted with GenAI-PA and achieved significantly better academic performance than the control group ($g = 0.428$). Besides, when Ruan et al. (2021) developed an English pedagogical agent for college students with mature cognitive development and strong autonomous learning ability, they found that the effect of GENAI-PA was mediocre ($g = 0.136$). To sum up, we cannot simply conclude that the influence of GenAI-PA on academic performance increases with the improvement of grade level, which is worthy of further analysis and demonstration.

Gender. Gender differences in technology self-efficacy may influence learners' acceptance of GenAI-PA and thereby impact academic performance. According to the socio-technical theory, males typically demonstrated higher confidence in using technology and greater willingness to adopt new digital tools, while females may experience higher levels of technological anxiety, which led to lower adoption rates (Cartelli, 2007). Empirical findings on the effect remain mixed. For example, Ngbarabara and Odibo (2024) found that male students in science education achieved significantly greater performance improvement than their female counterparts when taught with GenAI-PA. However, Wang et al. (2025) found no significant gender differences in the benefits derived from GenAI-PA. These conflicting findings suggest that gender may act as a conditional moderator mediating the effects of GenAI-PA on learners' academic performance. Therefore, further research is needed to explore the gender's moderating role on GenAI-PA effectiveness.

Instructional Characteristics

Learning Domain. Differences across subject areas may shape how effectively GenAI-PA supports student learning. Zhu et al. (2023) conducted a quasi-experimental study with 130 undergraduate students from formal sciences disciplines (e.g., mathematics, computer science) and humanity disciplines (e.g., literature, philosophy), using ChatGPT as an instructional tool. The results showed a greater effect on formal sciences students ($g = 0.316$) than on humanity students ($g = 0.078$). This is plausibly attributable to the systematic and rationally guided nature of formal sciences. This nature corresponds well with GenAI-PA's strengths in automated proof generation and real-time error correction, thus efficiently diminishing the learning curve (Zhuang, 2025). However, the role of GenAI-PA in the field of natural sciences (e.g., biology, physics, chemistry) has been questioned. Forero and Herrera-Suárez (2023) found a negative impact on physics learning ($g = -0.364$), as GenAI-PA only provided simple or incorrect answers, leading to a superficial understanding of physics knowledge. These mixed findings highlight the need for further analysis of the moderating effect of the learning domain.

Instructional Approach. Instructional approaches integrated with GenAI-PA can be classified into two types based on the degree of teacher involvement: self-directed learning (SDL) and teacher-directed learning (TDL) (Yasmin et al., 2019). SDL is characterized by students' autonomous engagement with GenAI-PA for knowledge construction through dynamic interaction (Ali et al., 2023). The effectiveness of this modality depends on learners' metacognitive ability and the intelligent systems' adaptive learning support function (Jin et al., 2023). TDL requires learners to have structured interactions with GenAI-PA under the educators' scaffolding and monitoring (Schweder et al., 2025). This approach emphasizes that GenAI-PA served as an instructional augmentation within teacher-led frameworks. Different instructional approaches may affect learners' performance when interacting with GenAI-PA. For instance, empirical evidence from Wei's (2023) quasi-experimental study demonstrated substantial improvement in college students' learning motivation and writing performance (g = 0.843) when GenAI-PA was integrated within teacher-facilitated scaffolding environments. In contrast, the randomized controlled trial by Lin and Chang (2020) showed that when students used GenAI-PA for SDL, the improvement in students' English composition scores was negligible (g = 0.039). These contrasting findings underscore the importance of examining the moderating effect of the instructional approach on the efficacy of GenAI-PA.

Learning Duration. Prior research also revealed the learning durations as a potential moderator in the effects of GenAI-PA on learners' academic performance. There are three possibilities for the moderating role of learning duration: First, GenAI-PA has a high effect size in short-term learning (0–4 weeks), as short-term intervention is more likely to stimulate students' learning motivation. For example, intelligent agents significantly improved students' healthcare skill levels (g = 0.780) through situational simulated conversations over two weeks (Lee et al., 2022). Second, GenAI-PA can enhance deep learning through dynamic planning during long-term learning (over 12 weeks), improving learners' academic performance significantly. An AI-powered math tutor provided six months of additional instruction to Ghanaian primary school students, significantly boosting math achievement (g = 0.451) (Henkel et al., 2024). Third, the appropriate length of learning duration (5–12 weeks) might ensure the optimal effect of GenAI-PA. Liu et al. (2022) used an intelligent reading agent to give personalized guidance to 39 primary school students for 7 weeks. Reasonable learning duration design avoided technical fatigue and achieved good results (g = 0.457). The above three inferences have not reached a unified consensus. Therefore, this study intends to explore the optimal learning duration of GenAI-PA to provide evidence to support further practice optimization.

GenAI-PA Characteristics

Instructional Role. Many researchers have noticed that the differences in the instructional roles simulated by PA can have different impacts on learners' academic performance (Baylor & Kim, 2005; Haake & Gulz, 2009; Tao et al., 2022). Among them, the widely applied instructional roles classification of PA was proposed by Baylor and Kim (2005). Based on the theory of social cognitive learning, they distinguished three instructional roles: Expert, Motivator, and Mentor (Kim & Baylor, 2016). The instructional role played by GenAI-PA can still follow this classification. The Expert agent typically builds authority by professional expressions and maintains an emotional distance from learners, focusing on conveying knowledge in a clear and concise way. For instance, Alneyadi and Wardat (2023) confirmed that GenAI-PA, playing the role of an expert, could very effectively enhance students' mastery of electromagnetism knowledge (g = 1.448). Nevertheless, the expert-centric unidirectional output model might potentially stifle innovative thinking among younger learners (Han et al., 2022). In contrast, the Motivator agent is often designed as a learning partner, with passionate and energetic conversations. The Motivator agent can encourage learners to persist in the task and stimulate learners to reflect by asking active questions. Kim (2019) found that using the Motivator agent for English conversation practice could significantly improve learners' oral expression level (g = 0.488), especially for students with high learning anxiety who need more emotional support. The Mentor agent integrates the characteristics of the Expert agent and the Motivator agent, providing both information and incentives. Winkler et al. (2020) used the Motivator agent in programming education. This type of GenAI-PA could not only explain concepts as an expert but also assisted in debugging code as a peer, significantly improving the learning effect (g = 0.689). In conclusion, GenAI-PA appears to benefit learning across different instructional role positions. However, whether the instructional role positioning mediates the impact of GenAI-PA on learners' academic performance warrants further investigation.

Dialogue Modality. The dialogue modality of GenAI-PA affects the efficiency of social cues transmission and thereby influences the academic performance of learners. Text dialogue is suitable for logical content, with the advantages of high information density and traceability, but may increase cognitive load (Rapp et al., 2021). Voice dialogue enhances emotional transmission through auditory channels and is suitable for language learning or younger students (Xu et al., 2021). Multimodal dialogue can balance the needs of efficiency and emotion and improve learners' learning experience through multi-sensory stimulation (Tai et al., 2024). Previous researchers have explored whether GenAI-PA with different dialogue modalities can have different effects on learners' academic performance. For example, Winkler et al. (2020) explored the impact of using text-dialogue GenAI-PA, voice-dialogue GenAI-PA, and multimodal dialogue GenAI-PA on students' learning programming. The results showed that the multimodal dialogue (g = 0.689) was more beneficial to student learning than a single text dialogue (g = 0.027) or voice dialogue (g = 0.174). However, whether this conclusion is reliable remains to be further explored through meta-analysis.

Method

This study used meta-analysis to examine the overall effectiveness of GenAI-PA in enhancing learners' academic performance and further investigated the moderating factors. Specifically, we estimated the pooled effect size of GenAI-PA on learners' academic performance and investigated how learner characteristics, instructional characteristics, and GenAI-PA characteristics moderated this result.

Literature Search and Screening

We searched the literature on the Web of Science, ERIC, ProQuest, IEEE Xplore, Google Scholar, SpringerLink, and EBSCOhost. The search keywords of GenAI-PA include Generative AI-powered Pedagogical Agents, Intelligent Pedagogical

Agents, Generative AI Agents, AI Tutor, Educational Conversational Agents, Educational AI Chatbot, etc. The search keywords of academic performance include Academic Performance, Academic Achievement, Academic Success, Learning Performance, Learning Outcome, etc. The search was limited to peer-reviewed journal articles published in English and papers from major international conferences with recognized academic standards (e.g., CHI conference on human factors in computing systems, International Conference on Artificial Intelligence in Education), from the first of January 2015 till the 31st of March 2025 to ensure the inclusion of the most recent and methodologically rigorous studies. To ensure the comprehensiveness of literature inclusion, in addition to database search, this study also used citation backtracking as a supplement to literature search. As shown in Figure 2, the initial search yielded 323 studies. After eliminating duplicates, 289 studies remained.

The inclusion criteria included: (1) The research was an empirical study to explore the impact of GenAI-PA on learners' academic performance. (2) The subjects were in K-12 or higher education. (3) The research object was GenAI-PA, excluding rule-based agents or script-driven agents. (4) The research design consisted of an experimental group (using GenAI-PA) and a control group (traditional teaching or no agent intervention). (5) The research reported quantitative data related to academic performance (e.g., mean, standard deviation, sample size) or extractable effect size. In the preliminary screening stage, a total of 176 non-empirical studies were excluded by title and abstract, which were unrelated to the research themes. A total of 86 studies with non-K-12 or non-higher education stage, rule-based agents or script-driven agents, non-controlled studies, and incomplete data were further excluded after full text review. Finally, 27 high-quality studies were included.

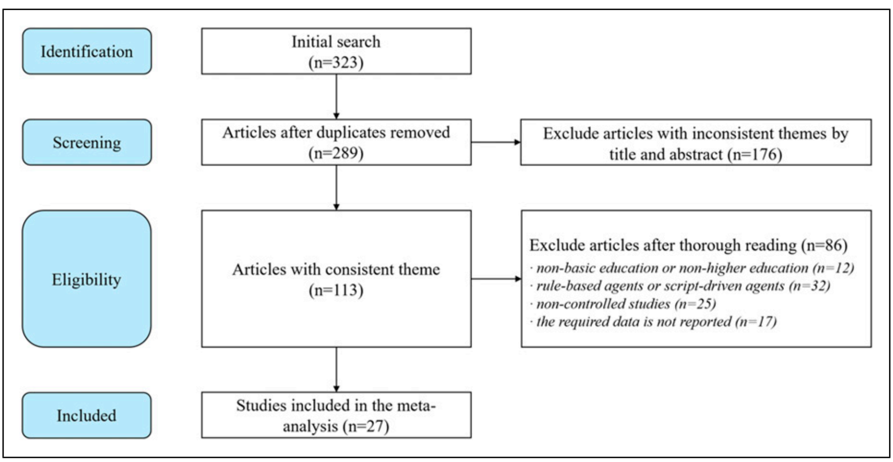

Figure 2. Flowchart for the Studies Selection Process

Data Coding

In order to achieve the research purpose, we used the content analysis approach recommended by Hsu et al. (2013) to extract relevant data information from the research literature: author, year of publication, total sample size, potential moderating variables (e.g., cultural background, grade level, gender, learning domain, instructional approach, learning duration, GenAI-PA's instructional role and GenAI-PA's dialogue modality), and effect size statistics (e.g., the sample sizes, means, and standard deviations of the experimental and control groups). Then, two researchers trained in coding independently conducted multiple rounds of coding by the following criteria: (1) To ensure the independence of the effect size, the effect size was coded separately for studies with only one independent sample; (2) if a study contains multiple independent samples, the effect sizes for each sample are coded separately, meaning that two or more effect sizes can be calculated in some studies. The inter-coder agreement reached 92%, with a Cohen's kappa of 0.87, indicating a high level of consistency between coders (Landis & Koch, 1977). Finally, we identified 31 independent effect sizes from the 27 articles that had been selected (see Table 1).

The specific coding scheme for the potential moderating variables is as follows: The cultural background was coded as collectivist culture (i.e., the individualism index of country/region is less than 50) and individualist culture (i.e., the individualism index of country/region is greater than or equal to 50). The individualism index data was derived from the website of Geert Hofstede. The grade level was coded as primary school (K-grade5), secondary school (grade 6-12), and university (college and higher vocational college). The proportion of female students was recorded when coding the gender. The learning domain was coded as humanities (e.g., literature, history, philosophy, linguistics, art), social sciences (e.g., sociology, psychology, economics, political science), natural sciences (e.g., physics, chemistry, biology, astronomy), formal sciences (e.g., mathematics, computer science, statistics, logic) (Löwe, 2002). The instructional approach was coded as self-directed learning and teacher-directed learning. The learning duration was coded as short-term learning (0 to 4 weeks), medium-term learning (5 to 12 weeks), and long-term learning (over 12 weeks). The instructional role of GenAI-PA was coded as expert, motivator, and mentor. The dialogue modality of GenAI-PA was coded as text, voice, or multimodal.

Data Analysis

This study employed Comprehensive Meta-Analysis 3.0 (CMA3.0) software for data analysis. The dataset consisted of 31 coded results derived from 27 valid studies, with GenAI-PA as the independent variable and learners' academic performance as the dependent variable. Cultural background, gender (female proportion), grade, learning domain, instructional approach, learning duration, instructional role and dialogue modality of GenAI-PA were treated as moderating variables. Hedges' *g* was utilized as an effect size to evaluate the influence of GenAI-PA on learners' academic performance. This standardized mean difference was chosen because it provides an unbiased

Table 1. Studies Included in the Meta-Analysis

Author (year)	Total sample size	Learner			Instruction			GenAI-PA	
		Grade level[a]	Female %[b]	Cultural background	Learning domain[c]	Instructional approach	Learning duration	Instructional role[d]	Dialogue modality[e]
Alneyadi and Wardat (2023)	122	2	0.475	Collectivist	3	TDL	Short	1	3
Chang et al.(2021)	36	3	N	Collectivist	2	TDL	Short	2	3
Cheon et al. (2025)	181	3	0.220	Collectivist	1	SDL	Short	1	2
Choi et al. (2024)	31	3	N	Collectivist	4	TDL	Short	1	2
Forero and Herrera-Suárez (2023)	76	3	N	Collectivist	3	SDL	Long	3	1
Fryer et al. (2017)	122	3	N	Collectivist	1	TDL	Medium	2	2
Guo et al. (2023)	44	3	0.545	Collectivist	1	SDL	Short	3	1
Han et al. (2022)	61	3	0.918	Collectivist	2	SDL	Short	1	1
Henkel et al. (2024)	477	1	N	Collectivist	4	TDL	Long	1	1
Hooshyar et al. (2015)	15	3	N	Collectivist	4	TDL	Short	1	1
Khazanchi et al. (2024)	78	2	0.460	Individualist	4	SDL	Medium	1	1
Kim (2019)	70	3	N	Collectivist	1	TDL	Long	2	1
Kumar (2021)	60	3	0.817	Collectivist	2	TDL	Long	1	3
Lee et al. (2022)	38	3	0.211	Collectivist	2	SDL	Short	1	1
Liu et al. (2022)	78	1	N	Collectivist	1	SDL	Medium	2	1
Mageira et al. (2022)	35	2	N	Individualist	1	TDL	Short	1	3
Manabete and Anyim (2024)	72	2	N	Collectivist	2	TDL	Short	1	3
Ngbarabara and Odibo (2024)	360	3	N	Collectivist	3	SDL	Medium	3	1
Pasang et al. (2024)	40	2	N	Collectivist	1	SDL	Short	3	3

(continued)

Table 1. (continued)

Author (year)	Total sample size	Learner			Instruction			GenAI-PA	
		Grade level[a]	Female %[b]	Cultural background	Learning domain[c]	Instructional approach	Learning duration	Instructional role[d]	Dialogue modality[e]
Lin and Chang (2020)	357	3	N	Individualist	1	SDL	Long	1	1
Ruan et al. (2021)	56	3	0.607	Collectivist	1	SDL	Short	3	3
Tai and Chen (2024)	62	2	N	Collectivist	1	SDL	Long	1	3
Wang et al. (2025)	12	3	0.583	Individualist	4	SDL	Short	2	3
Wei (2023)	60	3	0.600	Collectivist	2	TDL	Medium	3	3
Winkler et al. (2020) A	72	3	0.389	Individualist	4	SDL	Short	3	3
Winkler et al. (2020) B	72	3	0.361	Individualist	4	SDL	Short	3	1
Winkler et al. (2020) C	72	3	0.375	Individualist	4	SDL	Short	3	2
Xu et al. (2021) A	59	1	N	Individualist	1	TDL	Short	2	2
Xu et al. (2021) B	64	1	N	Individualist	1	TDL	Short	2	2
Zhu et al.(2023) A	96	3	N	Individualist	3	TDL	Short	3	1
Zhu et al.(2023) B	71	3	N	Individualist	2	TDL	Short	3	1

[a]For "Grade level": 1 = primary school, 2 = secondary school, 3 = university.
[b]For "Female%": N = not reported.
[c]For "Learning domain": 1 = humanities, 2 = social sciences, 3 = natural sciences, 4 = formal sciences.
[d]For "Instructional role": 1 = expert, 2 = motivator, 3 = mentor.
[e]For "Dialogue modality": 1 = text, 2 = voice, 3 = multimodal.

estimate of the effect size, particularly suitable for studies with small or unequal sample sizes (Hedges, 1992). Subpopulation analysis was conducted for categorical variables to explore differences in effect sizes across subpopulations. For the continuous variable, meta-regression analysis was performed under the assumption that the effect size may vary with changes in the continuous variable. This study aimed to construct a simple linear regression model to investigate the relationship between the effect size (Hedges' *g*) and continuous independent variables (female proportion). In terms of data analysis methods, this study adopted a heterogeneity test, main effect analysis, robustness test, moderator analysis, and meta-regression analysis to examine variations in the impact of GenAI-PA on learners' academic performance.

Results

This study included 27 experimental and quasi-experimental studies from 2015 to 2025, totaling 31 effect sizes, involving 3,049 students from 15 countries/regions, with individual study sample sizes ranging from 12 to 477.

Heterogeneity Test

To obtain more accurate and reliable combined effect sizes in meta-analysis, it is necessary to conduct a heterogeneity test on the included literature in the meta-analysis (Higgins & Thompson, 2002). A heterogeneity test refers to the examination of the degree of variation among each effect size in the meta-analysis, which is commonly assessed using Q statistic or I^2 statistic (Ruppar, 2020). When heterogeneity is present, a random-effects model should be employed; otherwise, a fixed-effects model should be used (Borenstein et al., 2010). According to the heterogeneity test results shown in Table 2, Q = 134.442 (p < .001), indicating the presence of heterogeneity among the included studies. Furthermore, according to the interpretation of I^2 by Huedo-Medina et al. (2006), I^2 = 77.686% (>75%) indicates a high level of heterogeneity. This suggests substantial variability among the results of the included studies, implying that the effects of GenAI-PA on learners' academic

Table 2. Homogeneity Test Results and Combined Effect Size of Random Effects Model

				Homogeneity test			Tau-squared			Test of null (two-tailed)	
Model	*N*	Hedges's g	95% CI	Q	*p*	I^2	Tau^2	SE	Tau	Z-value	*p*
REM	31	0.401	[0.236, 0.566]	134.442	.000	77.686	0.154	0.066	0.392	4.768	.000

**p* < .05, ***p* < .01, ****p* < .001.

performance may be strongly context-dependent. Therefore, moderator analyses were conducted to further explore the potential sources of heterogeneity. In view of the significant heterogeneity, this study adopted a random-effects model to provide a more appropriate estimation of the overall effect size and to account for between-study variance.

Effect Sizes

Based on the results of the heterogeneity test in Table 2, this study employed a random-effects model to conduct a main effect analysis of the impact of GenAI-PA on learners' academic performance, using Hedges' *g* as the standardized measure of effect size. Prior to the analysis, potential outliers were examined using standardized residual diagnostics. No studies showed standardized residuals exceeding ±3, which indicates that there were no significant outliers in this meta-analysis (Viechtbauer & Cheung, 2010). The main effect analysis results in Figure 3 indicated that the use of GenAI-PA significantly enhanced learners' academic performance compared to traditional learning (g = 0.401, $p < .001$). According to Hattie's (2008) interpretation of effect size, where 0.2, 0.4, and 0.6 are the thresholds for small, medium, and large effects, respectively. This suggested that GenAI-PA based learning had a medium positive effect on learners' academic performance compared to traditional learning.

Study	Statistics for each study						
	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
Alneyadi and Wardat (2023)	0.997	0.192	0.037	0.621	1.374	5.189	0.000
Chang et al.(2021)	1.448	0.374	0.140	0.714	2.182	3.867	0.000
Cheon et al. (2025)	0.118	0.150	0.023	-0.176	0.413	0.787	0.431
Choi et al. (2024)	1.200	0.390	0.152	0.435	1.965	3.074	0.002
Forero and Herrera-Suárez (2023)	-0.656	0.236	0.056	-1.119	-0.194	-2.783	0.005
Fryer et al. (2017)	-0.364	0.183	0.033	-0.721	-0.006	-1.992	0.046
Guo et al. (2023)	0.612	0.308	0.095	0.007	1.216	1.983	0.047
Han et al. (2022)	-0.126	0.256	0.066	-0.628	0.377	-0.490	0.624
Henkel et al. (2024)	0.451	0.093	0.009	0.269	0.633	4.865	0.000
Hooshyar et al. (2015)	0.858	0.357	0.127	0.159	1.557	2.407	0.016
Khazanchi et al. (2024)	0.147	0.228	0.052	-0.300	0.594	0.645	0.519
Kim (2019)	0.488	0.243	0.059	0.012	0.963	2.010	0.044
Kumar (2021)	0.978	0.273	0.075	0.443	1.514	3.581	0.000
Lee et al. (2022)	0.780	0.337	0.114	0.119	1.440	2.313	0.021
Liu et al. (2022)	0.457	0.259	0.067	-0.051	0.965	1.762	0.078
Mageira et al. (2022)	0.062	0.338	0.114	-0.602	0.725	0.182	0.856
Manabete and Anyim（ 2024）	1.929	0.280	0.078	1.380	2.478	6.890	0.000
Ngbarabara and Odibo (2024)	0.384	0.107	0.011	0.175	0.592	3.601	0.000
Pasang et al. (2024)	0.796	0.329	0.108	0.152	1.440	2.423	0.015
Pin-Chuan Lin, and Chang (2020)	0.039	0.106	0.011	-0.169	0.247	0.365	0.715
Ruan et al. (2021)	0.136	0.268	0.072	-0.388	0.660	0.508	0.611
Tai and Chen (2024)	0.428	0.257	0.066	-0.075	0.932	1.666	0.096
Wang et al. (2025)	-0.415	0.413	0.170	-1.224	0.394	-1.006	0.314
Wei (2023)	0.843	0.269	0.073	0.315	1.372	3.131	0.002
Winkler et al. (2020)A	0.689	0.243	0.059	0.213	1.165	2.839	0.005
Winkler et al. (2020)B	0.027	0.236	0.056	-0.435	0.489	0.115	0.909
Winkler et al. (2020)C	0.174	0.236	0.056	-0.289	0.637	0.738	0.460
Xu et al. (2021)A	0.435	0.265	0.070	-0.085	0.955	1.639	0.101
Xu et al. (2021)B	-0.024	0.250	0.063	-0.514	0.466	-0.096	0.923
Zhu et al. (2023)A	0.316	0.228	0.052	-0.131	0.764	1.387	0.166
Zhu et al. (2023)B	0.078	0.260	0.068	-0.432	0.588	0.299	0.765
	0.401	0.084	0.007	0.236	0.566	4.768	0.000

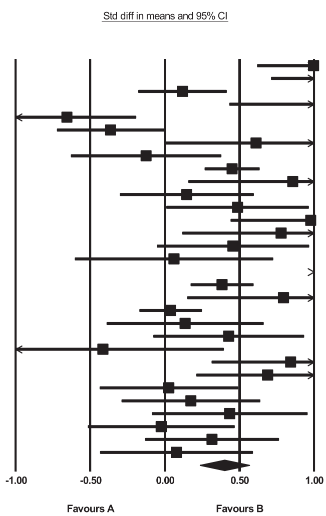

Figure 3. The Main Effect Analysis Results

Moderator Analysis

To further explore the differences in the impact of GenAI-PA on learners' academic performance under various moderating variables, this study conducted moderation effect tests on learner characteristics (i.e., cultural background, grade level, gender), instructional characteristics (i.e., learning domain, instructional approach, learning duration), and GenAI-PA characteristics (i.e., instructional role, dialogue modality) (see Tables 3 and 4).

Learner Characteristics. Cultural background significantly moderated the impact of GenAI-PA on learners' academic performance (Q = 8.956, $p < .01$). Specifically, the effect size of collectivist culture was 0.547 (k = 20), and that of individualist culture was 0.141 (k = 11), indicating that the use of GenAI-PA in the context of collectivist culture was more likely to improve learners' academic performance.

In terms of grade level, the analysis results showed that grade level did not moderate the impact of GenAI-PA on learners' academic performance (Q = 2.113, $p > .05$). GenAI-PA had a large effect on middle school ($g = 0.731$, $k = 6$), and only a lower moderate effect on primary school ($g = 0.393$, $k = 4$) and university ($g = 0.318$, $k = 21$). Similarly, there was no gender difference in GenAI-PA in improving learners' academic performance (SE = 0.569, Z = −0.07, $p > .05$).

Instructional Characteristics. The instructional approach indicated a significant moderating effect (Q = 5.770, $p < .05$). Specifically, GenAI-PA was more effective in teacher-directed learning contexts ($g = 0.616$, $k = 15$) than in self-directed learning contexts ($g = 0.211$, $k = 16$).

No significant results found in learning domains (Q = 4.928, $p > .05$) and durations (Q = 1.479, $p > .05$). Nevertheless, the effect sizes varied across subject domains: the social sciences exhibited the largest effect size ($g = 0.834$, $k = 7$), followed by the formal sciences ($g = 0.376$, $k = 8$), the natural sciences ($g = 0.275$, $k = 4$), and the humanities ($g = 0.209$, $k = 12$). Regarding learning duration, GenAI-PA showed the greatest effectiveness in short-term learning ($g = 0.489$, $k = 20$), whereas both medium-term ($g = 0.271$, $k = 5$) and long-term ($g = 0.274$, $k = 6$) learning yielded relatively lower effect sizes.

GenAI-PA Characteristics. Dialogue modality revealed a significant moderating effect (Q = 7.007, $p < .01$). Specifically, GenAI-PA employing multimodal dialogue (e.g., combining text, voice, and visual elements) showed the highest effect size ($g = 737$, $k = 11$), whereas those using text-only ($g = 0.248$, $k = 14$) or voice-only dialogues ($g = 0.176$, $k = 6$) demonstrated substantially lower effectiveness.

On the other hand, the instructional role settings of GenAI-PA did not significantly moderate its impact on learners' academic performance (Q = 2.581, $p > .05$). Nevertheless, the effect sizes varied: the expert role exhibited the highest effect size ($g = 0.568$, $k = 13$), followed by the mentor role ($g = 0.292$, $k = 11$) and the motivator role

Table 3. Results of Moderator Analysis

Moderating variables	$Q_{\text{between-groups}}$	k	Hedges's g (95% CI)	Heterogeneity test
Learner characteristics				
Cultural background	Q = 8.956**			
Collectivist culture		20	0.547 [0.317, 0.778]***	Q = 109.794***, I^2 = 82.695
Individualist culture		11	0.141 [0.009, 0.274]*	Q = 10.462, I^2 = 4.413
Grade levels	Q = 2.113			
Primary School		4	0.393 [0.219, 0.567]***	Q = 3.243, I^2 = 7.480
Secondary School		6	0.731 [0.201, 1.262]**	Q = 32.115***, I^2 = 84.431
University		21	0.318 [0.126, 0.509]**	Q = 77.800***, I^2 = 74.293
Instructional characteristics				
Learning domain	Q = 4.928			
Humanities		12	0.209 [0.031, 0.387]*	Q = 21.384*, I^2 = 48.559
Social Sciences		7	0.834 [0.281, 1.387]**	Q = 40.361***, I^2 = 85.134
Natural Science		4	0.275 [−0.286, 0.835]	Q = 29.654***, I^2 = 89.883
Formal Science		8	0.376 [0.123, 0.629]**	Q = 16.117*, I^2 = 56.568
Instructional approach	Q = 5.770*			
Self-directed learning		16	0.211 [0.042, 0.380]*	Q = 37.125, I^2 = 59.596
Teacher-directed learning		15	0.616 [0.332, 0.901]***	Q = 80.131, I^2 = 82.529
Learning duration	Q = 1.479			
Short-term		20	0.489 [0.249, 0.728]***	Q = 78.961***, I^2 = 75.938
Medium-term		5	0.271 [−0.098, 0.640]	Q = 18.523**, I^2 = 78.406
Long-term		6	0.274 [−0.072, 0.620]	Q = 31.905***, I^2 = 84.328
GenAI-PA characteristics				
Instructional role	Q = 2.581			
Expert		13	0.568 [0.293, 0.843]***	Q = 72.949***, I^2 = 83.550
Motivator		7	0.270 [−0.149, 0.689]	Q = 26.744***, I^2 = 77.565
Mentor		11	0.292 [0.056, 0.528]**	Q = 29.736**, I^2 = 66.371
Dialogue modality	Q = 7.007**			
Text		14	0.248 [0.072, 0.423]**	Q = 37.015***, I^2 = 64.879
Voice		6	0.176 [−0.155, 0.507]	Q = 16.317**, I^2 = 69.357
Multimodal		11	0.737 [0.392, 1.081]***	Q = 42.600***, I^2 = 76.526

*$p < .05$, **$p < .01$, ***$p < .001$.

($g = 0.270$, $k = 7$). This result suggested that GenAI-PA configured as an expert tended to be more effective in supporting knowledge and skill acquisition.

Publication Bias

To ensure the accuracy and reliability of the meta-analysis results, it is necessary to conduct robustness tests in the process of meta-analysis (Fragkos, 2014), including publication bias and fail-safe number (Nfs). Publication bias refers to the fact that statistically significant studies are more likely to be accepted and published than non-statistically significant studies in the same class, or that high-response studies are more likely to be published than low-effect studies (Thornton & Lee, 2000). If there is no systematic difference between the missing literature and the included literature, there will be no systematic impact on the estimate of the effect size. Common methods for publishing bias tests include the funnel plot and Egger's test. It can be intuitively seen from the funnel plot (see Figure 4) that the 31 effector scatter points included in the study were distributed evenly and symmetrically on both sides of the standard line, indicating that there was no publication bias in this study. Further Egger regression test (see Table 5) showed that $t = 1.217$ ($p > .05$), which also proved that no publication bias was found in this study.

The fail-safe number (Nfs) was used to test the reliability of the results. When the Nfs is less than 5k + 10 (k is the original number of studies included in the meta-analysis), the possibility of publication bias is significant (Rothstein et al., 2005). Finally, the Nfs = 671 (>5*27 + 10) was much larger than the critical value, indicating that the results of this study have high reliability.

Discussion

This study aimed to determine the promoting effect of GenAI-PA on learners' academic performance and the moderating effect of related variables. Overall, the results showed that GenAI-PA can effectively enhance learners' academic performance, with this effect significantly moderated by cultural background, instructional approach, and dialogue modality. The following sections will discuss findings from the meta-analysis and provide recommendations in education practice.

Table 4. Meta-Regression Analyses

	COEF	SE	95% CI	Z	*p* (two-tailed)
Female (%)					
Intercept	0.430	0.289	[−0.137, 0.996]	1.49	0.137
Slope	−0.041	0.569	[−1.157, 1.075]	−0.07	0.942

$*p < .05$, $**p < .01$, $***p < .001$.

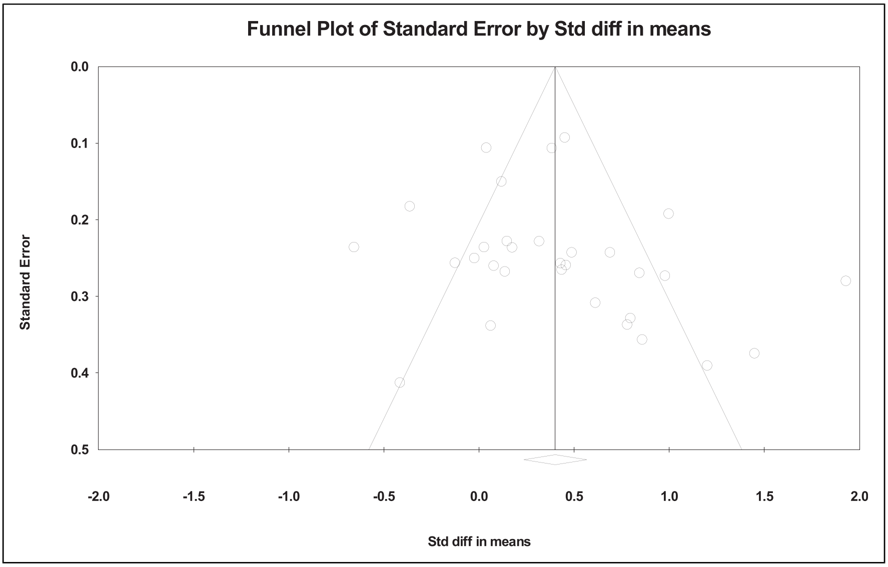

Figure 4. Funnel Plot of Standard Error by Effect Sizes

Main Effect: The Promoting Effect of GenAI-PA on Academic Performance

Through meta-analysis, this study found that the application of GenAI-PA can significantly improve learners' academic performance (g = 0.401, p < .001). This conclusion echoes previous meta-analysis studies on PA (Castro-Alonso et al., 2021; Guo et al., 2015; Wang et al., 2017), but GenAI-PA showed a higher main effect size. This result can be attributed to the following three aspects. First, dynamic generation ability optimizes personalized learning paths. PA mostly relied on preset rules and limited databases, and the feedback form was fixed (Winkler et al., 2020). However, GenAI-PA, powered by LMMs, analyzes the semantic context of students' input in real time. It uses deep learning trained on a massive corpus to generate dynamic and adaptive feedback (McGuire et al., 2024). For example, in programming problem solving, GenAI-PA can not only point out calculation errors, but also guide students to independently correct thinking loopholes through multi-step reasoning thinking chain questioning (Wang et al., 2025). This iterative mechanism of "generation-correction-regeneration" significantly strengthens learners' metacognitive ability (Kasneci et al.,

Table 5. Egger's Regression Test Results

	Intercept	SE	95% CI	t	*p* (two-tailed)
Academic performance	1.123	0.923	[−0.764, 3.010]	1.217	0.233

*p < .05, **p < .01, ***p < .001.

2023). Second, contextualized dialogue enhances social learning experiences. GenAI-PA can simulate the coherence and emotional temperature of real teacher-student dialogue by relying on the strong context understanding ability of LMMs. Studies have shown that when GenAI-PA adopted “Socratic questioning” or “empathic encouragement”, students were more likely to establish a sense of trust in the machine and thus improve their learning engagement (Park & Kim, 2023). In language learning, GenAI-PA can automatically generate revision suggestions of various styles based on students’ compositions and create immersive practice situations through role-playing (such as “editing” and “debating opponents”) to make up for the monotony and lag of traditional writing feedback (Tai & Chen, 2024). Third, facilitating deep learning with multimodal interactions. GenAI-PA breaks through the limitations of single text interaction and can integrate text, image, voice, video, and other multimodal content generations. For example, students can use natural language instructions to have GenAI-PA generate pictures, text of complex concepts, and deepen conceptual understanding by incorporating speech explanations (Wei, 2023). It is important to note that the benefits of GenAI-PA are predicated on high-quality training data and ethical constraints. Some studies have shown that LMMs may produce “hallucination” due to data biases, such as generating false facts or contradictory logic (Elsayed, 2024). Therefore, it is necessary to establish a manual audit mechanism in the application of GenAI-PA in education. GenAI-PA needs to be positioned as a collaborator rather than a replacer. Ultimately, technical efficiency can be achieved through a three-way collaboration among teachers, GenAI-PA, and students.

Learner: Collectivist Culture Enhances the Effectiveness of GenAI-PA

The study found that the effect size of GenAI-PA in the context of collectivist culture ($g = 0.547$) was significantly higher than that in the context of individualist culture ($g = 0.141$). This finding echoes sociocultural theory, which emphasizes the interplay between social interaction and cognitive development, indicating that cultural values have different effects on learning behavior (Marginson & Dang, 2016). Collectivist culture emphasizes collaboration and authority compliance (Hofstede, 2011). GenAI-PA, which aims to facilitate peer discussion or simulate collaborative problem-solving, may be aligned with the values of collectivist shared knowledge construction. Thus, students are more adapted to the structured learning process dominated by GenAI-PA with the assistance of teachers. In contrast, individualist culture attaches importance to self-directed learning and critical thinking, and students may conflict with the fixed interaction mode of GenAI-PA, resulting in limited technical utility (Sun et al., 2024). Therefore, future GenAI-PA development should prioritize cultural adaptation, embedding features that echo local educational practices, whether through language localization, specific content, or alignment with regional curriculum goals.

Although grade level had no significant moderating effect, the effect size was higher in middle school than in primary school and university, and this difference could be explained by the characteristics of cognitive development and the adaptation of technology. Middle school students are in the formal operational stage of Piaget’s theory of cognitive development (Huitt & Hummel, 2003), and their abstract thinking and logical reasoning abilities develop rapidly. GenAI-PA's interactive and visualization support is what they need to represent abstract transformations. For primary school students in the concrete computing stage, GenAI-PA mainly relies on screen interaction and offers limited multi-sensory experiences. Because the interface operation often exceeds their technical proficiency, its overall effect tends to be mediocre (Xu et al., 2021). For undergraduates, the learning goal shifts to higher-order cognition and professional deepening. At present, GenAI-PA mainly focuses on basic knowledge transfer and standardized training, which makes it difficult to support the achievement of higher-order learning goals, resulting in limited instrumental value for university students (Escalante et al., 2023). Finally, gender has no moderating effect on the effect of GenAI-PA, indicating that the effect of GenAI-PA is universal. GenAI-PA reduces gender bias in traditional classrooms and provides equal learning opportunities for students of different genders (Halaweh, 2023).

Instruction: Human-Machine Collaboration is the Key to GenAI-PA Effectiveness

Among the three instructional characteristics, only the instructional approach had a significant moderating effect. The effect of GenAI-PA in teacher-directed learning was significantly better than self-directed learning, which confirmed the core value of human-machine collaboration (Ansari et al., 2018). Therefore, to maximize the learning benefits of GenAI-PA, it is essential to establish a human–machine collaboration model with a clear and balanced division of instructional responsibilities. As cognitive scaffolding, GenAI-PA undertakes standardized training (e.g., grammar correction, formula derivation) and instant question answering, freeing teachers' energy to lead higher-order thinking (e.g., critical discussion and metacognitive reflection) (Kim et al., 2022). On one hand, teachers, acting as facilitators, assist students in comprehending the logic behind GenAI-PA feedback, prevent the misuse of technology, and bolster their awareness of critical technology evaluation. For example, when students over-rely on the generation of answers, teachers can provide timely intervention and guidance (Felix, 2020). On the other hand, the integration of human-machine collaboration results through classroom discussion promotes knowledge transfer. For example, in collaborative learning, teachers can organize students to compare GenAI-PA solutions with peer perspectives to deepen conceptual understanding (Manabete & Anyim, 2024).

As for learning domains, the social science field had the highest effect size, but the difference between disciplines was not significant. Social science knowledge usually relies on contextualized discussion and multiple perspectives. Through case database matching and causal chain deduction, GenAI-PA could achieve limited generalization of case deduction and situational dialogue, stimulating students' critical thinking (Zhu et al., 2023). The deployment of GenAI-PA in formal science and natural science education proved effective. However, it risks becoming merely an answer assistant for students, potentially undermining the deep knowledge construction of some learners

(Khazanchi et al., 2024). The minimal effect size in the humanities might stem from GenAI-PA's affective computing being restricted to semantic emotion recognition, failing to fulfill the humanistic teaching need for emotional resonance and empathic dialogue (Yang et al., 2021).

In terms of learning duration, the effect size of short-term learning was higher than that of medium and long-term learning, but the difference was not significant. The possible reasons lie in the novelty effect of GenAI-PA, which can stimulate students' learning interest at the initial stage, whereas its repetitive and standardized interaction patterns may lead to cognitive fatigue after prolonged use (Li & Liu, 2022). In addition, some studies failed to control the increasing difficulty of the learning domain, resulting in inadequate adaptability of GenAI-PA in long-term intervention (Wu & Yu, 2024). Future research needs to design a dynamic adjustment mechanism so that GenAI-PA can optimize the feedback strategy along with the learning process.

GenAI-PA: Multimodal Dialogue has Greater Educational Potential

The results indicated that GenAI-PA with multimodal dialogue outperforms text and voice dialogue. This aligns with the cognitive theory of multimedia learning, where multichannel information input enhances working memory integration (Mayer, 2005). For instance, GenAI-PA combined with phonetic correction and text annotation can simultaneously enhance listening and speaking ability in language learning (Cheon et al., 2025). The low effect size of pure speech interaction may be due to its transient nature and lack of visual anchoring, resulting in low information retention (Winkler et al., 2020). Therefore, developers should prioritize multimodal design, leveraging advances in Natural Language Processing and GenAI to create immersive, context-aware interactions.

In the instructional roles of GenAI-PA, the expert role had the highest GenAI-PA effect size, but the instructional role had no significant moderating effect. The authority of the Expert agents may enhance learners' sense of trust, especially in knowledge transfer (Wang et al., 2023). However, the difference of roles effect may also be affected by application scenarios: in skill training, the step-by-step guidance of the Expert agents is more effective; in creative tasks, the Motivator agents are more advantageous (Kuhail et al., 2023). Future research should explore role adaptability, for example, agents that switch between Expert and Motivator roles based on task complexity or student confidence level.

These findings confirm the positive impact of GenAI-PA on learners' academic performance, and reveal that this effect is moderated by cultural background, instructional approach, and dialogue modality. Rather than serving as automated feedback systems, GenAI-PA operate as adaptive learning partners whose performance is shaped by sociocultural and instructional conditions. The moderating role of cultural background underscores the importance of culturally adaptive GenAI-PA design that aligns with local educational philosophies, instructional goals, and communication norms. The stronger effects observed in teacher-directed learning underscore the importance of human–AI collaboration, where GenAI-PA provide cognitive scaffolding while teachers guide higher-order reasoning. In addition, the advantages of multimodal dialogue extend the cognitive theory of multimedia learning by demonstrating how multimodal interaction supports cognitive integration. Collectively, these insights contribute to a more comprehensive theoretical basis for the development and application of GenAI-PA, and offer a coherent foundation for designing culturally adaptive, multimodally enriched, and human–AI collaborative learning environments.

Conclusion and Limitations

Conclusion

Through meta-analysis, this study confirmed that GenAI-PA had a moderate positive effect on learners' academic performance, and its effectiveness was significantly moderated by cultural background, instructional approach, and dialogue modality. Based on this, we put forward the following three suggestions on the design and application of GenAI-PA.

Focus on Culturally Adaptive Design to Achieve Localization and Contextualization. The development of GenAI-PA needs to be combined with specific cultural contexts to avoid "one-size-fits-all" technology transplantation (Barnes et al., 2024). Specifically, designers should deeply analyze the cultural characteristics of target users (e.g., language habits, educational norms, social values) and build an interactive framework that conforms to local cognitive logic. For example, it is necessary to strengthen the open dialogue ability of GenAI-PA and support learners' independent exploration and demonstration in the Western educational environment. In addition, multi-language models and cross-cultural corpus training can be introduced to improve GenAI-PA's ability to understand cultural metaphors, idioms, and social situations, to avoid the attenuation of learning efficiency caused by semantic misunderstanding (Orosoo et al., 2024).

Customized Development for Different Instructional Activities From the Perspective of Human-Machine Collaboration. Firstly, according to the characteristics of the learning activities, such as the goal, content, duration, and environment, teachers should build a professional knowledge base of the discipline, design the GenAI-PA function, and even achieve "one customized GenAI-PA for one activity" (Zhang et al., 2025). Secondly, it is necessary to establish a learning ecology of "human-machine collaboration" to avoid the loss of human interaction caused by over-reliance on technology. For example, in blended learning scenarios, GenAI-PA can be positioned as an "intelligent collaborator" that helps teachers monitor group progress and provide real-time feedback, rather than completely replacing the teacher's dominant role (Järvelä et al., 2023).

Deepen the Multimodal Dialogue Modality While Avoiding Cognitive Load. Research shows that multimodal dialogue GenAI-PA has the strongest effect on learning, which indicates that comprehensive interactive modes can meet the diversified needs of learners and improve the efficiency of information transfer. But voice-only dialogue is worse than text-only dialogue, it is a warning that inappropriate multimodal designs can lead to cognitive overload. Therefore, the dialogue modality design of GenAI-PA should follow the principle of appropriateness (Yuan, 2024). In low complexity tasks such as data search and fact memory, text-based and visually assisted, lightweight interaction should be adopted. In higher-order tasks such as problem-solving, design, and reasoning simulation, voice and animation can be integrated to enhance situational cognition through immersive experience.

Limitations

This study followed strict standards and procedures for conducting a meta-analysis. It systematically reviewed existing studies on the effectiveness of GenAI-PA in improving learners' academic performance and thereby produced more objective conclusions. However, it still has limitations. Due to the limited number of original studies, the number of effect values among subgroups is unevenly distributed for the two moderating variables of grade level and learning duration, which may lead to potential bias in the analysis and interpretation of these two types of variables. Given the above deficiencies, future studies should further expand the coverage of research objects, enrich the number of studies on different cultural backgrounds, learning domains, etc. At the same time, due to the emergence and rapid development of new technologies, the function and form of GenAI-PA are constantly reshaped and upgraded. Future research should follow the pace of scientific and technological development and continue to explore the actual utility of GenAI-PA in the context of new technologies.

ORCID iDs

Liang Cheng https://orcid.org/0009-0008-9858-1806
Hui Shi https://orcid.org/0000-0002-7954-5841
Yuhan Wu https://orcid.org/0009-0009-5122-7532
Feng Li https://orcid.org/0000-0001-9182-6075

Author Contributions

Liang Cheng: Conceptualization; Methodology; Data curation; Formal analysis; Writing - original draft preparation; Writing – review & editing. Hui Shi: Conceptualization; Writing - original draft preparation; Writing – review & editing. Yuhan Wu: Data curation; Writing – review & editing. Feng Li: Conceptualization; Writing - original draft preparation; Writing – review & editing; Supervision.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Declaration of Conflicting Interests

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

Data Availability Statement

The data and code supporting the findings of this study are available upon request.

References

Ali, F., Choy, D., Divaharan, S., Tay, H. Y., & Chen, W. (2023). Supporting self-directed learning and self-assessment using teacher GAIA, a generative AI chatbot application: Learning approaches and prompt engineering. *Learning: Research and Practice*, *9*(2), 135–147. https://doi.org/10.1080/23735082.2023.2258886

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

Ansari, F., Erol, S., & Sihn, W. (2018). Rethinking human-machine learning in industry 4.0: How does the paradigm shift treat the role of human learning? *Procedia Manufacturing*, *23*(1), 117–122. https://doi.org/10.1016/j.promfg.2018.04.003

Apoki, U. C., Hussein, A. M. A., Al-Chalabi, H. K. M., Badica, C., & Mocanu, M. L. (2022). The role of pedagogical agents in personalised adaptive learning: A review. *Sustainability*, *14*(11), 6442. https://doi.org/10.3390/su14116442

Arguedas, M., Daradoumis, T., & Caballé, S. (2024). Measuring the effects of pedagogical agent cognitive and affective feedback on students' academic performance. *Frontiers in Artificial Intelligence*, *7*(1), 1495342. https://doi.org/10.3389/frai.2024.1495342

Barnes, A. J., Zhang, Y., & Valenzuela, A. (2024). AI and culture: Culturally dependent responses to AI systems. *Current Opinion in Psychology*, *58*(1), 101838. https://doi.org/10.1016/j.copsyc.2024.101838

Baylor, A. L., & Kim, Y. (2005). Simulating instructional roles through pedagogical agents. *International Journal of Artificial Intelligence in Education*, *15*(2), 95–115. https://doi.org/10.3233/IRG-2005-15(2)02

Bektik, D., Ullmann, T. D., Edwards, C., Herodotou, C., & Whitelock, D. (2024). AI-Powered curricula: Unpacking the potential and progress of generative technologies in education. *Ubiquity Proceedings*, *4*(1), 38. https://doi.org/10.5334/uproc.160

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, *13*(6), 4–16. https://doi.org/10.3102/0013189x013006004

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods*, *1*(2), 97–111. https://doi.org/10.1002/jrsm.12

Borges, J. M., & Araújo, R. D. (2024). Experiences and challenges of a redesign process with the support of an AI assistant on an educational platform. In *Proceedings of the XXIII Brazilian symposium on human factors in computing systems* (pp. 1–12). ACM. https://doi.org/10.1145/3702038.3702046

Bozkurt, A. (2023). Generative artificial intelligence (AI) powered conversational educational agents: The inevitable paradigm shift. *Asian Journal of Distance Education*, *18*(1), 198–214. https://doi.org/10.5281/ZENODO.7716416

Cartelli, A. (2007). Socio-technical theory and knowledge construction: Towards new pedagogical paradigms? *Issues in Informing Science and Information Technology*, *4*(1), 1–14. https://doi.org/10.28945/928

Castro-Alonso, J. C., Wong, R. M., Adesope, O. O., & Paas, F. (2021). Effectiveness of multimedia pedagogical agents predicted by diverse theories: A meta-analysis. *Educational Psychology Review*, *33*(3), 989–1015. https://doi.org/10.1007/s10648-020-09587-1

Chang, C. Y., Hwang, G. J., & Gau, M. L. (2021). Promoting students' learning achievement and self-efficacy: A mobile chatbot approach for nursing training. *British Journal of Educational Technology*, *53*(1), 171–188. https://doi.org/10.1111/bjet.13158

Cheon, G., Choi, Y., Lee, D., & Baek, J. (2025). Generative AI agents in language learning: A randomized field experiment. In *Proceedings of the 58th Hawaii International Conference on System Sciences* (pp. 4949-4958). https://hdl.handle.net/10125/109443

Choi, S., Kang, S., Lee, K., Ju, H., & Song, J. (2024). The effect of an agent tutor's integration of cognitive and emotional gestures on cognitive load, motivation, and achievement. *Contemporary Educational Technology*, *16*(1), ep491. https://doi.org/10.30935/cedtech/14101

Chugh, R., Morshed, M., Sabrina, F., Azad, S., Rashid, M. M., Kaisar, K., Subramani, S., Kutty, S., & Turnbull, D. (2025). *AI-powered pedagogy: Generative AI as a learning assistant in the ICT discipline*. CQ University. https://doi.org/10.25946/28561211.v1

Dai, Y., Xiao, J. Y., Huang, Y., Zhai, X., Wai, F. C., & Zhang, M. (2025). How generative AI enables an online project-based learning platform: An applied study of learning behavior analysis in undergraduate students. *Applied Sciences*, *15*(5), 2369. https://doi.org/10.3390/app15052369

Domagk, S. (2010). Do pedagogical agents facilitate learner motivation and learning outcomes? The role of the appeal of agent's appearance and voice. *Journal of Media Psychology: Theories, Methods, and Applications*, *22*(2), 84–97. https://doi.org/10.1027/1864-1105/a000011

Efklides, A. (2006). Metacognition and affect: What can metacognitive experiences tell us about the learning process? *Educational Research Review*, *1*(1), 3–14. https://doi.org/10.1016/j.edurev.2005.11.001

Elsayed, H. (2024). The impact of hallucinated information in large language models on student learning outcomes: A critical examination of misinformation risks in AI-assisted education. *Northern Reviews on Algorithmic Research, Theoretical Computation, and Complexity*, *9*(8), 11–23. https://northernreviews.com/index.php/NRATCC/article/view/2024-08-07

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

Felix, C. V. (2020). The role of the teacher and AI in education. In. *International perspectives on the role of technology in humanizing higher education* (pp. 33–48). Emerald Publishing Limited. https://doi.org/10.1108/S2055-364120200000033003

Forero, M. G., & Herrera-Suárez, H. J. (2023). ChatGPT in the Classroom: Boon or Bane for Physics Students' Academic Performance? *arXiv preprint arXiv:2312.02422*. https://doi.org/10.48550/ARXIV.2312.02422

Fragkos, K. C., Tsagris, M., & Frangos, C. C. (2014). Publication bias in meta-analysis: Confidence intervals for Rosenthal's fail-safe number. *International Scholarly Research Notices*, *2014*(1), 825383. https://doi.org/10.1155/2014/825383

Fryer, L. K., Ainley, M., Thompson, A., Gibson, A., & Sherlock, Z. (2017). Stimulating and sustaining interest in a language course: An experimental comparison of chatbot and human task partners. *Computers in Human Behavior*, *75*(1), 461–468. https://doi.org/10.1016/j.chb.2017.05.045

Guo, K., Zhong, Y., Li, D., & Chu, S. K. W. (2023). Effects of chatbot-assisted in-class debates on students' argumentation skills and task motivation. *Computers & Education*, *203*(1), 104862. https://doi.org/10.1016/j.compedu.2023.104862

Guo, Y. R., & Goh, D. H.-L. (2015). Affect in embodied pedagogical agents: Meta-analytic review. *Journal of Educational Computing Research*, *53*(1), 124–149. https://doi.org/10.1177/0735633115588774

Haake, M., & Gulz, A. (2009). A look at the roles of look & roles in embodied pedagogical agents–a user preference perspective. *International Journal of Artificial Intelligence in Education*, *19*(1), 39–71. https://doi.org/10.3233/IRG-2009-19(1)03

Halaweh, M. (2023). ChatGPT in education: Strategies for responsible implementation. *Contemporary Educational Technology*, *15*(2), ep421. https://doi.org/10.30935/cedtech/13036

Han, J. W., Park, J., & Lee, H. (2022). Analysis of the effect of an artificial intelligence chatbot educational program on non-face-to-face classes: A quasi-experimental study. *BMC Medical Education*, *22*(1), 830. https://doi.org/10.1186/s12909-022-03898-3

Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement* (Vol. *59*). Routledge. https://doi.org/10.4324/9780203887332

Hedges, L. V. (1992). Meta-analysis. *Journal of Educational Statistics*, *17*(4), 279–296. https://doi.org/10.3102/10769986017004279

Heilman, M., Collins-Thompson, K., Callan, J., & Eskenazi, M. (2006). Classroom success of an intelligent tutoring system for lexical practice and reading comprehension. In *Proceedings of the ninth international conference on spoken language processing*. IEEE. https://doi.org/10.21437/Interspeech.2006-282

Henkel, O., Horne-Robinson, H., Kozhakhmetova, N., & Lee, A. (2024). Effective and scalable math support: Experimental evidence on the impact of an AI-Math tutor in Ghana. In *International conference on artificial intelligence in education* (pp. 373–381). Springer. https://doi.org/10.1007/978-3-031-64315-6_34

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

Hofstede, G. (2011). Dimensionalizing cultures: The Hofstede model in context. *Online Readings in Psychology and Culture*, *2*(1), 8. https://doi.org/10.9707/2307-0919.1014

Hooshyar, D., Ahmad, R. B., Yousefi, M., Yusop, F. D., & Horng, S.- (2015). A flowchart-based intelligent tutoring system for improving problem-solving skills of novice programmers. *Journal of Computer Assisted Learning*, *31*(4), 345–361. https://doi.org/10.1111/jcal.12099

Hsu, Y. C., Hung, J. L., & Ching, Y. H. (2013). Trends of educational technology research: More than a decade of international research in six SSCI-indexed refereed journals. *Educational Technology Research & Development*, *61*(4), 685–705. https://doi.org/10.1007/s11423-013-9290-9

Huedo-Medina, T. B., Sánchez-Meca, J., Marín-Martínez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I^2 index? *Psychological Methods*, *11*(2), 193–206. https://doi.org/10.1037/1082-989X.11.2.193

Huitt, W., & Hummel, J. (2003). Piaget's theory of cognitive development. In *Educational psychology interactive*. Valdosta State University. https://www.edpsycinteractive.org/topics/cognition/piaget.html

Järvelä, S., Nguyen, A., & Hadwin, A. (2023). Human and artificial intelligence collaboration for socially shared regulation in learning. *British Journal of Educational Technology*, *54*(5), 1057–1076. https://doi.org/10.1111/bjet.13325

Jin, S. H., Im, K., Yoo, M., Roll, I., & Seo, K. (2023). Supporting students' self-regulated learning in online learning using artificial intelligence applications. *International Journal of Educational Technology in Higher Education*, *20*(1), 37. https://doi.org/10.1186/s41239-023-00406-5

Joshi, S. (2025). The transformative role of agentic GenAI in shaping workforce development and education in the US. *Iconic Research and Engineering Journals*, *8*(8), 199–206. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5133376

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, *103*(1), 1–9. https://doi.org/10.1016/j.lindif.2023.102274

Khazanchi, R., Di Mitri, D., & Drachsler, H. (2025). The effect of AI-based systems on mathematics achievement in rural context: A quantitative study. *Journal of Computer Assisted Learning*, *41*(1), e13098. https://doi.org/10.1111/jcal.13098

Kim, M. K., Kim, N. J., & Heidari, A. (2022). Learner experience in artificial intelligence-scaffolded argumentation. *Assessment & Evaluation in Higher Education*, *47*(8), 1301–1316. https://doi.org/10.1080/02602938.2022.2042792

Kim, N. Y. (2019). A study on the use of artificial intelligence chatbots for improving English grammar skills. *Journal of Digital Convergence*, *17*(8), 37–46. https://doi.org/10.14400/JDC.2019.17.8.037

Kim, Y., & Baylor, A. L. (2016). Research-based design of pedagogical agent roles: A review, progress, and recommendations. *International Journal of Artificial Intelligence in Education*, *26*(1), 160–169. https://doi.org/10.1007/s40593-015-0055-y

Kuhail, M. A., Alturki, N., Alramlawi, S., & Alhejori, K. (2023). Interacting with educational chatbots: A systematic review. *Education and Information Technologies*, *28*(1), 973–1018. https://doi.org/10.1007/s10639-022-11177-3

Kumar, J. A. (2021). Educational chatbots for project-based learning: Investigating learning outcomes for a team-based design course. *International Journal of Educational Technology in Higher Education*, *18*(1), 65. https://doi.org/10.1186/s41239-021-00302-w

Landis, J. R., & Koch, G. G. (1977). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics*, *33*(2), 363–374. https://doi.org/10.2307/2529786

Lee, Y. F., Hwang, G. J., & Chen, P. Y. (2022). Impacts of an AI-based chabot on college students' after-class review, academic performance, self-efficacy, learning attitude, and motivation. *Educational Technology Research & Development*, *70*(5), 1843–1865. https://doi.org/10.1007/s11423-022-10142-8

Li, X., & Liu, D. (2022). The influence of technostress on cyberslacking of college students in technology-enhanced learning: Mediating effects of deficient self-control and burnout. *International Journal of Environmental Research and Public Health*, *19*(18), 11800. https://doi.org/10.3390/ijerph191811800

Lin, M. P. C., & Chang, D. (2020). Enhancing post-secondary writers' writing skills with a chatbot. *Journal of Educational Technology & Society*, *23*(1), 78–92. https://www.jstor.org/stable/26915408

Lin, Y., & Yu, Z. (2025). Learner perceptions of artificial intelligence-generated pedagogical agents in language learning videos: Embodiment effects on technology acceptance. *International Journal of Human-Computer Interaction*, *41*(2), 1606–1627. https://doi.org/10.1080/10447318.2024.2359222

Liu, C. C., Liao, M. G., Chang, C. H., & Lin, H. M. (2022). An analysis of children' interaction with an AI chatbot and its impact on their interest in reading. *Computers & Education*, *189*(1), 104576. https://doi.org/10.1016/j.compedu.2022.104576

Löwe, B. (2002). The formal sciences: Their scope, their foundations, and their unity. *Synthese*, *133*(1/2), 5–11. https://www.jstor.org/stable/20117290

Mageira, K., Pittou, D., Papasalouros, A., Kotis, K., Zangogianni, P., & Daradoumis, A. (2022). Educational AI chatbots for content and language integrated learning. *Applied Sciences*, *12*(7), 3239. https://doi.org/10.3390/app12073239

Makransky, G., Wismer, P., & Mayer, R. E. (2019). A gender matching effect in learning with pedagogical agents in an immersive virtual reality science simulation. *Journal of Computer Assisted Learning*, *35*(3), 349–358. https://doi.org/10.1111/jcal.12335

Manabete, S. S., & Anyim, C. A. (2024). Effects of artificial intelligence on senior secondary two civic education students' motivation in Jos-North local government area, plateau state. *BW Academic Journal*. https://bwjournal.org/index.php/bsjournal/article/view/2509

Marginson, S., & Dang, T. K. A. (2016). Vygotsky's sociocultural theory in the context of globalization. *Asia Pacific Journal of Education*, *37*(1), 116–129. https://doi.org/10.1080/02188791.2016.1216827

Matsuda, N., Cohen, W. W., & Koedinger, K. R. (2015). Teaching the teacher: Tutoring SimStudent leads to more effective cognitive tutor authoring. *International Journal of Artificial Intelligence in Education*, *25*(1), 1–34. https://doi.org/10.1007/s40593-014-0020-1

Mayer, R. E. (2005). Cognitive theory of multimedia learning. *The Cambridge Handbook of Multimedia Learning*, *41*(1), 57–72. https://doi.org/10.1017/9781108894333.008

Mayer, R. E., Sobko, K., & Mautone, P. D. (2003). Social cues in multimedia learning: Role of speaker's voice. *Journal of Educational Psychology*, *95*(2), 419–425. https://doi.org/10.1037/0022-0663.95.2.419

McGuire, A., Qureshi, W., & Saad, M. (2024). A constructivist model for leveraging GenAI tools for individualized, peer-simulated feedback on student writing. *International Journal of Technology in Education*, *7*(2), 326–352. https://doi.org/10.46328/ijte.639

Mehdian, N., & Turi, A. N. (2025). AI-Powered educational system: Challenges for higher education in emerging economies: Baseline research. *Tech Transformation and AI Readiness*, *113*(1), 113–125. https://doi.org/10.1007/978-3-031-73639-1_7

Mohammed, I. A., Bello, A., & Ayuba, B. (2025). Effect of large language models artificial intelligence ChatGPT chatbot on achievement of computer education students. *Education and Information Technologies*, *30*(9), 11863–11888. https://doi.org/10.1007/s10639-024-13293-8

Mousavinasab, E., Zarifsanaiey, N., Niakan Kalhori, S., Rakhshan, M., Keikha, L., & Ghazi Saeedi, M. (2021). Intelligent tutoring systems: A systematic review of characteristics, applications, and evaluation methods. *Interactive Learning Environments*, *29*(1), 142–163. https://doi.org/10.1080/10494820.2018.1558257

Ngbarabara, P. B., & Odibo, A. (2024). Effect of generative artificial intelligence (GAI) on science education students' achievements in Federal University Otuoke, Bayelsa State. *FUO-Journal of Educational Research*, *3*(2), 121–131. https://doi.org/10.5281/zenodo.13825732

Orosoo, M., Goswami, I., Alphonse, F. R., Fatma, G., Rengarajan, M., & Bala, B. K. (2024, March). *Enhancing natural language processing in multilingual chatbots for cross-cultural communication*. In *2024 5th International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)* (pp. 127-133). IEEE. https://doi.org/10.1109/ICICV62344.2024.00027

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. *The Internet and Higher Education*, *65*(1), 100991. https://doi.org/10.1016/j.iheduc.2024.100991

Park, D. Y., & Kim, H. (2023). Determinants of intentions to use digital mental healthcare content among university students, faculty, and staff: Motivation, perceived usefulness, perceived ease of use, and parasocial interaction with AI chatbot. *Sustainability*, *15*(1), 872. https://doi.org/10.3390/su15010872

Pasang, S., Latifa, A., & Ammade, S. (2024). The use of kuki chatbot application to improve English achievement. *International Journal of Education and Literacy Studies*, *6*(1), 15–46. https://doi.org/10.32996/ijels.2024.6.1.3

Rapp, A., Curti, L., & Boldi, A. (2021). The human side of human-chatbot interaction: A systematic literature review of ten years of research on text-based chatbots. *International Journal of Human-Computer Studies*, *151*(1), 102630. https://doi.org/10.1016/j.ijhcs.2021.102630

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005). Publication bias in meta-analysis. In *Publication bias in meta-analysis: Prevention, assessment and adjustments* (pp. 1–7). Wiley. https://doi.org/10.1002/0470870168.ch1

Ruan, S., Jiang, L., Xu, Q., Liu, Z., Davis, G. M., Brunskill, E., & Landay, J. A. (2021). EnglishBot: An AI-powered conversational system for second language learning. In *Proceedings of the 26th international conference on intelligent user interfaces* (pp. 434–444). ACM. https://doi.org/10.1145/3397481.3450648

Ruiz Viruel, S., Sánchez Rivas, E., & Ruiz Palmero, J. (2025). The role of artificial intelligence in project-based learning: Teacher perceptions and pedagogical implications. *Education Sciences*, *15*(2), 150. https://doi.org/10.3390/educsci15020150

Ruppar, T. (2020). Meta-analysis: How to quantify and explain heterogeneity? *European Journal of Cardiovascular Nursing*, *19*(7), 646–652. https://doi.org/10.1177/1474515120944014

Schneider, S., Beege, M., Nebel, S., Schnaubert, L., & Rey, G. D. (2022). The cognitive-affective-social theory of learning in digital environments (CASTLE). *Educational Psychology Review*, *34*(1), 1–38. https://doi.org/10.1007/s10648-021-09626-5

Schroeder, N. L., Davis, R. O., & Yang, E. (2024). Designing and learning with pedagogical agents: An umbrella review. *Journal of Educational Computing Research*, *62*(8), 1907–1936. https://doi.org/10.1177/07356331241288476

Schweder, S., Grahl, L., & Raufelder, D. (2025). Examining self-directed and teacher-directed learning's impact on achievement goals and learning strategies. *The Journal of Experimental Education*, 1–26. https://doi.org/10.1080/00220973.2025.2459382

Shetye, S. (2024). An evaluation of Khanmigo, a generative AI tool, as a computer-assisted language learning app. *Studies in Applied Linguistics and TESOL*, *24*(1), 38–53. https://doi.org/10.52214/salt.v24i1.12869

Sun, J., Wang, Y., & Qian, Z. (2024). Examining the influence of individual-level cultural values on CFL learners' acceptance of ChatGPT for Chinese learning. *Interactive Learning Environments*, *33*(5), 3393–3407. https://doi.org/10.1080/10494820.2024.2443785

Tai, T. Y., & Chen, H. H. J. (2024). The impact of intelligent personal assistants on adolescent EFL learners' listening comprehension. *Computer Assisted Language Learning*, *37*(3), 433–460. https://doi.org/10.1080/09588221.2022.2040536

Tao, Y., Zhang, G., Zhang, D., Wang, F., Zhou, Y., & Xu, T. (2022). Exploring persona characteristics in learning: A review study of pedagogical agents. *Procedia Computer Science*, *201*(1), 87–94. https://doi.org/10.1016/j.procs.2022.03.014

Thornton, A., & Lee, P. (2000). Publication bias in meta-analysis: Its causes and consequences. *Journal of Clinical Epidemiology*, *53*(2), 207–216. https://doi.org/10.1016/S0895-4356(99)00161-4

Viechtbauer, W., & Cheung, M. W. L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods*, *1*(2), 112–125. https://doi.org/10.1002/jrsm.11

Wang, F., Li, W., Xie, H., & Liu, H. (2017). Is pedagogical agent in multimedia learning good for learning? A meta-analysis. *Advances in Psychological Science*, *25*(1), 12. https://doi.org/10.3724/SP.J.1042.2017.00012

Wang, T., Wu, T., Liu, H., Brown, C., & Chen, Y. (2025). Generative co-learners: Enhancing cognitive and social presence of students in asynchronous learning with generative AI. *Proceedings of the ACM on Human-Computer Interaction*, *9*(1), 1–24. https://doi.org/10.1145/3701198

Wei, L. (2023). Artificial intelligence in language instruction: Impact on English learning achievement, L2 motivation, and self-regulated learning. *Frontiers in Psychology*, *14*(1), 1261955. https://doi.org/10.3389/fpsyg.2023.1261955

Wei, X., Wang, L., Lee, L.-K., & Liu, R. (2024). Multiple generative AI pedagogical agents in augmented reality environments: A study on implementing the 5E model in science education. *Journal of Educational Computing Research*, *63*(2), 336–371. https://doi.org/10.1177/07356331241305519

Winkler, R., Hobert, S., Salovaara, A., Söllner, M., & Leimeister, J. M. (2020). Sara, the lecturer: Improving learning in online education with a scaffolding-based conversational agent. In *Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1–14). ACM. https://doi.org/10.1145/3313831.3376781

Wu, R., & Yu, Z. (2024). Do AI chatbots improve students' learning outcomes? Evidence from a meta-analysis. *British Journal of Educational Technology*, *55*(1), 10–33. https://doi.org/10.1111/bjet.13334

Xu, Y., Wang, D., Collins, P., Lee, H., & Warschauer, M. (2021). Same benefits, different communication patterns: Comparing children's reading with a conversational agent vs. a human partner. *Computers & Education*, *161*(1), 104059. https://doi.org/10.1016/j.compedu.2020.104059

Yan, L., Greiff, S., Teuber, Z., & Gašević, D. (2024). Promises and challenges of generative artificial intelligence for human learning. *Nature Human Behaviour*, *8*(10), 1839-1850. https://doi.org/10.1038/s41562-024-02004-5

Yang, S. J., Ogata, H., Matsui, T., & Chen, N. S. (2021). Human-centered artificial intelligence in education: Seeing the invisible through the visible. *Computers and Education: Artificial Intelligence*, *2*(1), 100008. https://doi.org/10.1016/j.caeai.2021.100008

Yasmin, M., Naseem, F., & Masso, I. C. (2019). Teacher-directed learning to self-directed learning transition barriers in Pakistan. *Studies in Educational Evaluation*, *61*(1), 34–40. https://doi.org/10.1016/j.stueduc.2019.02.003

Yuan, B. (2024). Design of an intelligent dialogue system based on natural language processing. *Journal of Theory and Practice of Engineering Science*, *4*(01), 72–78. https://doi.org/10.53469/jtpes.2024.04(01).10

Żemojtel-Piotrowska, M., & Piotrowski, J. (2023). Hofstede's cultural dimensions theory. In T. K. Shackelford (Ed.), *Encyclopedia of sexual psychology and behavior* (pp. 1–4). Springer International Publishing. https://doi.org/10.1007/978-3-031-08956-5_1124-1

Zhang, L., Yao, Z., & Hadizadeh Moghaddam, A. (2025). Designing GenAI tools for personalized learning implementation: Theoretical analysis and prototype of a multi-agent system. *Journal of Teacher Education*, *76*(3), 280–293. https://doi.org/10.1177/00224871251325109

Zhang, S., Jaldi, C. D., Schroeder, N. L., López, A. A., Gladstone, J. R., & Heidig, S. (2024). Pedagogical agent design for K-12 education: A systematic review. *Computers & Education*, *223*(1), 105165. https://doi.org/10.1016/j.compedu.2024.105165

Zhu, G., Fan, X., Hou, C., Zhong, T., Seow, P., Shen-Hsing, A. C., Rajalingam, P., Yew, L. K., & Poh, T. L. (2023). Embrace opportunities and face challenges: Using ChatGPT in undergraduate students' collaborative interdisciplinary learning. arXiv. https://doi.org/10.48550/arXiv.2305.18616

Zhuang, Y. (2025). Lessons from using ChatGPT in calculus: Insights from two contrasting cases. *Journal of Formative Design in Learning*, *9*(1), 25–35. https://doi.org/10.1007/s41686-025-00098-2

Author Biographies

Liang Cheng is a doctoral candidate in Department of Education Information Technology at East China Normal University. His research focuses on developing and applying artificial intelligence technology to promote teaching and learning.

Hui Shi is an assistant professor in the Faculty of Artificial Intelligence in Education at Central China Normal University. Her research focuses on using learning analytics and education data mining to support online learning participation and collaboration.

Yuhan Wu is a master in Department of Education Information Technology at East China Normal University. Her research interests center on using various effective strategies to carry out information technology education.

Feng Li is a Professor of Department of Education Information Technology at East China Normal University. His research focuses on designing and developing curriculum standards and teaching materials for artificial intelligence education.