THE EFFECTIVENESS OF GENERATIVE AI IN EDUCATION: A SYSTEMATIC REVIEW OF EMPIRICAL STUDY

Wang Xiaoyu[1]
***Chin Hai Leng**[1]
Zamzami Zainuddin[2]
[1] Department of Curriculum and Instructional Technology, Faculty of Education, Universiti Malaya, Malaysia
[2] College of Education, Psychology and Social Work, Flinders University, Australia
* *chin@um.edu.my*

Abstract: Generative AI has gained attention for its potential to transform education through personalized learning and improved outcomes. However, some institutions and educators have banned generative AI, viewing it as a Pandora's box. This review evaluates the effectiveness of generative AI in education and identifies factors influencing its implementation. A systematic review of 13 studies published between 2022 and 2024 was conducted using PRISMA guidelines. Seventy per cent of studies reported positive impacts on learning outcomes, particularly through personalized feedback. However, challenges hindered critical thinking and creativity when over-relied upon. Generative AI has potential, but its effectiveness depends on various factors. Future research should address concerns about creativity and the appropriate use of AI while exploring diverse educational contexts and methods.

Keywords: Generative AI, Education, Learning Outcomes, Systematic Review, PRISMA, Educational Technology

INTRODUCTION

Generative Artificial Intelligence (AI), recognized for its ability to mimic human creativity and intelligence, facilitates the generation of various media types (text, images, and videos) through platforms like ChatGPT and other Large Language Models (LLMs). Unlike traditional AI models that primarily analyze and respond to inputs, generative AI models can produce original outputs, which has led to a wide array of applications across diverse domains (Sanhita Kar et al., 2023). Since its emergence at the end of 2022, generative AI has demonstrated excellent performance and has been widely adopted across fields, including engineering, healthcare, finance, and education (Bahroun et al., 2023). In particular, the education sector has increasingly embraced generative AI, exploring its potential to transform teaching and learning processes, from generating course materials to automating responses to student queries (Blagoev et al., 2023; Kadaruddin, 2023).

While the adoption of generative AI in education is growing, the academic community remains divided on its effectiveness in enhancing teaching and learning outcomes (Ogunleye et al., 2024). Some studies suggest that generative AI positively impacts student learning by creating educational content, improving engagement, and personalizing learning experiences (Hakiki et al., 2023; Kasneci et al., 2023; Nguyen Thanh et al., 2023; Shahzad et al., 2024; Ray, 2023). Conversely, other studies raise concerns that generative AI may potentially harm student performance, stemming from inaccuracies, bias, misuse, and over-reliance (Fui-Hoon Nah et al., 2023; Kurtz et al., 2024; Zhu et al., 2024). Due to these risks, some institutions and educators have banned generative AI, likening it to a Pandora's box, especially given the lack of direct evidence showing a consistently positive impact on student performance (Ming & Mansor, 2023). Moreover, the empirical evidence on its impact remains fragmented and inconclusive, with a notable absence of research exploring generative AI's impact on enhancing student learning outcomes (Zhou & Kim, 2024). This evidence highlights the urgent need for further research to understand how generative AI influences educational outcomes across various contexts.

Given these discrepancies and the growing adoption of AI technologies in educational settings, this study seeks to address these gaps by systematically reviewing empirical studies on the effectiveness of generative AI in education. By synthesizing existing literature, this review aims to provide a comprehensive understanding of the patterns and trends in generative AI research, assess the effectiveness of generative AI in improving educational outcomes, and identify the factors that contribute to or hinder its success. The findings of this review would offer educators and policymakers evidence-based insights necessary for making informed decisions about integrating AI into curricula, thereby avoiding suboptimal outcomes and seizing opportunities to enhance learning.

RESEARCH PURPOSE, OBJECTIVES AND RESEARCH QUESTIONS

This systematic review explores the research findings concerning the relationship between generative AI use and academic performance in an educational context. This exploration examines various dimensions such as research sites, publication sources, methodological approaches, educational levels, and the specific generative AI platforms and apps employed. Understanding these dimensions is essential to uncover the current research's scope, focus, and methodologies, which would reveal trends, gaps, and potential biases in the literature. Additionally, the study aims to evaluate how generative AI impacts educational outcomes and to identify the factors that influence its successful implementation, providing insights into the conditions necessary for its optimal use in educational settings.

The objectives of this study are as follows:

1. To analyze the patterns of selected studies on generative AI in education, focusing on research sites, publication sources, methodological approaches, educational levels, and generative AI platforms and apps.
2. To evaluate the effective of generative AI in enhancing specific educational outcomes.
3. To identify the factors influencing the effect of generative AI in educational implementation.

The research questions that guide this study are:

1. What are the patterns of the reviewed studies in terms of research site, publication sources, methodological approaches, educational levels, and generative AI platforms and apps?
2. How effective is generative AI in enhancing specific educational outcomes?
3. What factors influence the effect of Generative AI in educational implementation?

METHODS

Research Design

This research involved a systematic literature review to examine patterns and trends in generative AI research, assess its effectiveness in improving educational outcomes, and identify factors contributing or hindering its implementation. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines were followed to ensure transparency and methodological rigor. PRISMA was chosen for its ability to define the research question clearly, set inclusion/exclusion criteria, establish a timeframe for examining large scientific databases, and enable coding for future reviews (Sierra-Correa & Kintz, 2015).

Resources

The review for this study was conducted using a primary database, with academic journals selected from Clarivate Analytics' Web of Science to ensure the inclusion of high-quality, impactful scientific content. Web of Science is widely regarded as one of the most trusted citation indices for evidence-based, high-quality scientific information (Martín-Martín et al., 2018).

The Systematic Review Process

The systematic review process for selecting relevant articles in this study involved three key phases. Initially, keywords were identified to guide the search. Next, articles were screened according to predefined inclusion and exclusion criteria set by the researchers. Lastly, the eligibility of the remaining articles was assessed to finalize their inclusion.

1. Identification of sources

The process of identifying keywords for the search started by looking for related and synonymous terms, guided by previous studies, particularly Montenegro-Rueda et al. (2023) and Valverde-Berrocoso et al. (2022). Once the relevant keywords were established, search strings were created in the Web of Science database (see Table 1). In the initial phase of the systematic review, 159 articles were identified and retrieved from the database.

Table 1.
The Search Strings

Database	Search Strings
WoS	TS="generative AI*"or "generative artificial intelligence*"or "GenAI*"or "Large language model*"or "ChatGPT*" AND "learning outcomes*"or "academic

	achievement*" or "student performance*"or "educational outcomes*"or "student success*"

2. Screening

Next, 159 articles were screened based on several inclusion and exclusion criteria designed by the researchers. The inclusion criteria to identify the published articles for this review study are as follows: (1) the document had to be classified as an "Article," (2) it must have been published between 2022 and 2024, (3) the publication language was restricted to English, (4) the research had to focus on "Education Educational Research," (5) the articles had to be indexed in the Social Sciences Citation Index (SSCI), (6) they must have been published by reputable publishers such as Springer Nature, Elsevier, Taylor & Francis, Sage, Wiley, or Emerald Group Publishing, (7) the studies needed to be empirical, employing qualitative, quantitative, or mixed methods, (8) the research had to address the effectiveness of Generative AI in teaching and learning, and (9) the articles were required to include clear details about the countries involved, publication sources, research aims, methodologies, educational levels, and the GenAI platforms and apps used. In the end, 146 articles were excluded based on these criteria (see Table 2).

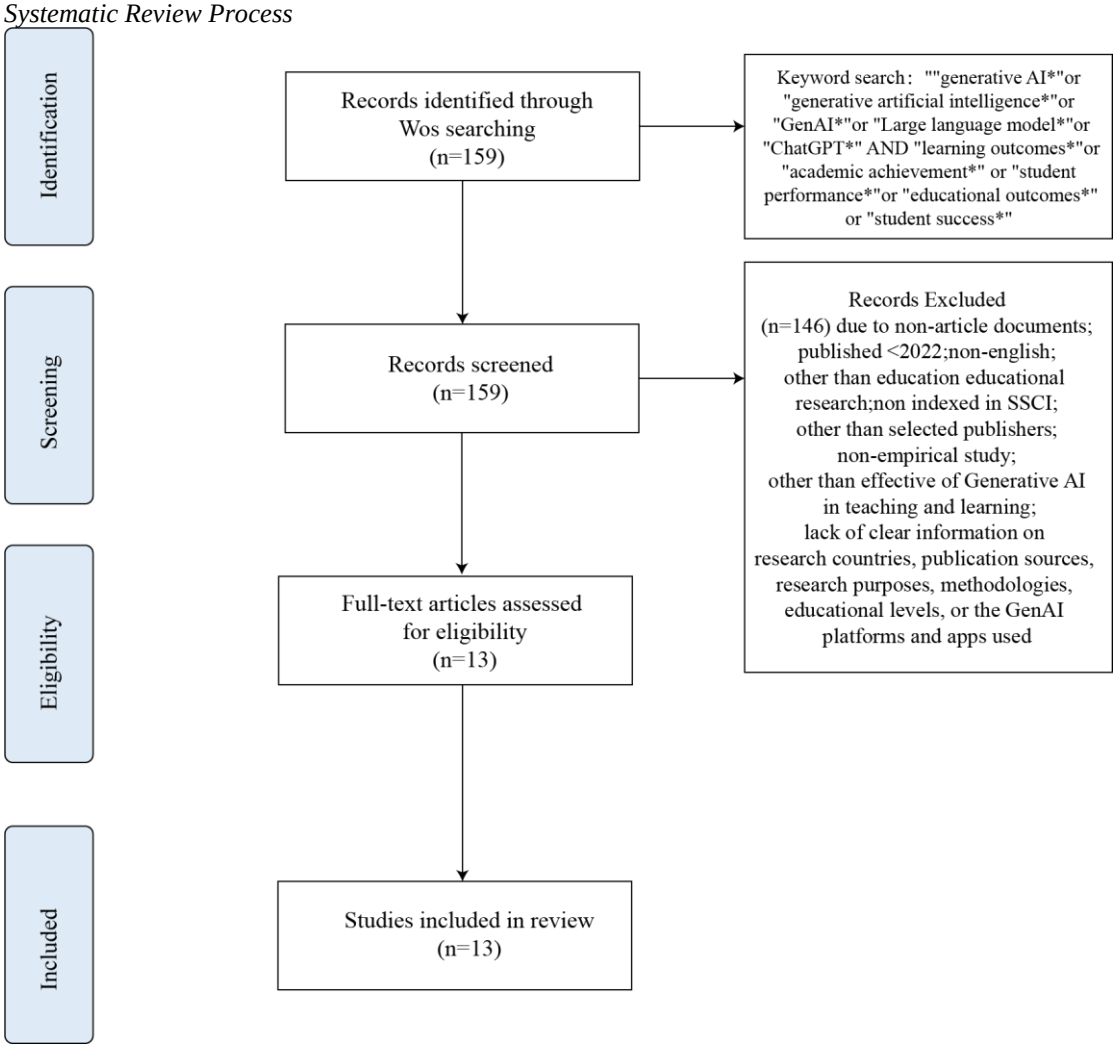
Table 2.
Criteria for Inclusion and Exclusion

Criterion	Eligibility	Exclusion
Document Type	Article documents	Non-article documents (e.g., reviews, editorials, or conference papers)
Timeline	Published between 2022 and 2024	Published outside the 2022–2024 timeframe
Language	English	Non-English publications
Research Areas	Education Educational Research	Research outside the field of "Education Educational Research"
Index	Social Sciences Citation Index (SSCI)	Not indexed in the Social Sciences Citation Index (SSCI)
Publisher	Springer Nature, Elsevier, Taylor & Francis, Sage, Wiley, or Emerald Group Publishing	Published by publishers other than Springer Nature, Elsevier, Taylor & Francis, Sage, Wiley, or Emerald Group Publishing
Type of study	Empirical studies (qualitative, quantitative, or mixed methods)	Non-empirical studies (i.e., studies that are not qualitative, quantitative, or mixed methods)
Research direction	Effective of Generative AI in teaching and learning	Research not focused on the effective of Generative AI in teaching and learning
Content of research	Clear information on research countries, publication sources, research purposes, methodologies, educational levels, or the GenAI platforms and tools used.	Lack of clear information on research countries, publication sources, research purposes, methodologies, educational levels, or the GenAI platforms and apps used.

3. Eligibility and inclusion

Subsequently, 13 articles were selected for a detailed eligibility assessment. Each article's title, abstract, research questions, problem statement, methodology, and findings were carefully reviewed to ensure they met the inclusion criteria and aligned with the objectives. A sheet was created in Microsoft Word to organize data extracted from the articles, such as research countries, publication sources, research aims, methodologies, educational levels, and the Generative AI platforms and apps used (See appendix). Ultimately, all 13 articles were fully analyzed. Figure 1 illustrates the systematic review process.

Figure 1.
Systematic Review Process



Data Analysis

An integrative review approach was employed to analyze the selected articles, allowing for the combination of various research methodologies, including qualitative, quantitative, and mixed methods. This approach offered the flexibility to transform qualitative data into a quantitative format or vice versa, ensuring a robust and thorough analysis (Whittemore & Knafl, 2005). The articles were reviewed not only to extract essential data but also to capture broader insights into generative AI's effectiveness in education

The first step involved extracting key information from each article, such as research countries, publication sources, research aims, methodologies, educational levels, and the Generative AI platforms and apps, all linked to the study's overarching research questions. This information was systematically organized to facilitate the development of themes. Through careful examination, three core themes emerged. Firstly, the patterns in the reviewed studies. Secondly, the effectiveness of generative AI in enhancing educational outcomes. Lastly, the

critical factors affecting its successful implementation. This thematic organization helped connect the diverse insights from the reviewed studies, revealing patterns that would otherwise remain unnoticed.

FINDINGS

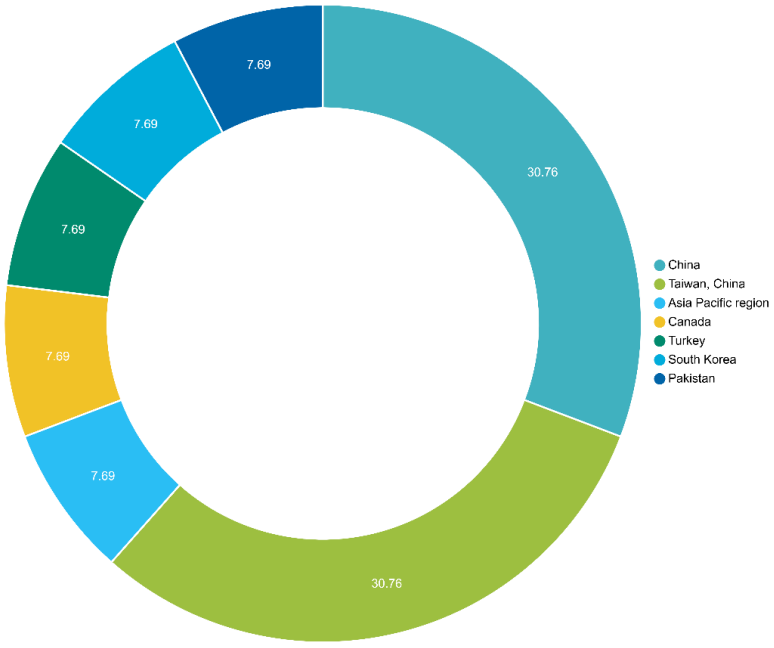
Patterns of Reviewed Studies (RQ1)

In this part of the study, we outline the results, concentrating on the recurring themes and trends observed in the reviewed studies to offer a thorough summary of the existing literature on the application of generative AI in education.

1. Research sites

As shown in Figure 2, the research reviewed spans seven countries or regions. A substantial portion of the studies was conducted in China (4 studies, 30.76%) and Taiwan, China (4 studies, 30.76%), followed by individual studies from the Asia Pacific region, Canada, Turkey, South Korea, and Pakistan (1 study each, 7.69%).

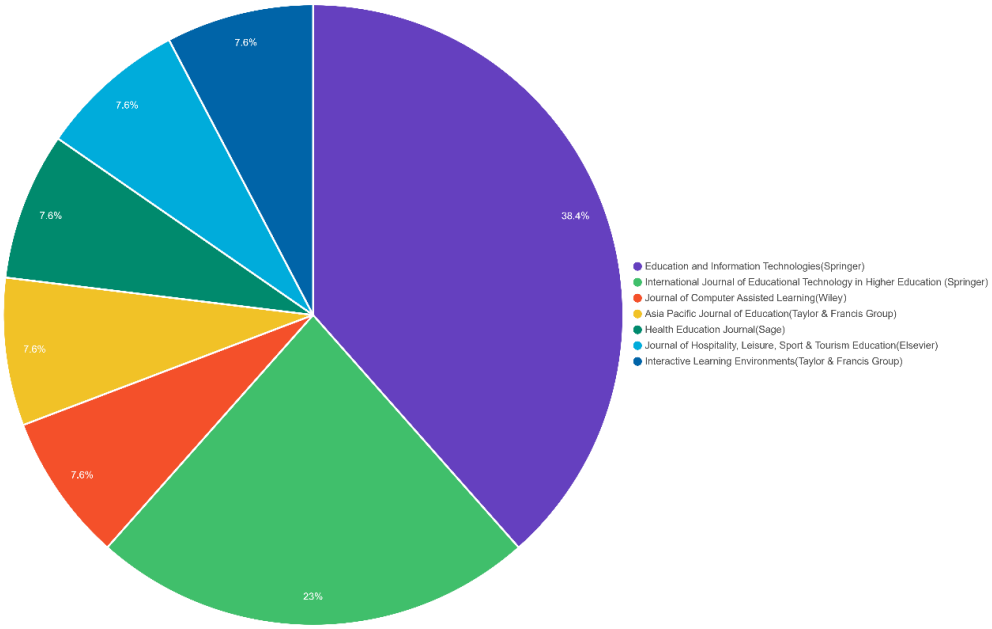
Figure 2.
Research Sites of Selected Studies



2. Publication sources

As illustrated in Figure 3, the studies in this review are distributed across seven different journals. A notable portion was published in “Education and Information Technologies” (5 studies, 38.46%) and followed by “International Journal of Educational Technology in Higher Education” by Springer (3 studies, 23.08%). Other contributions came from “Journal of Computer Assisted Learning” by Wiley (1 study, 7.69%), “Asia Pacific Journal of Education” by Taylor & Francis Group (1 study, 7.69%), “Health Education Journal” by Sage (1 study, 7.69%), “Journal of Hospitality, Leisure, Sport & Tourism Education” by Elsevier (1 study, 7.69%), and “Interactive Learning Environments” by Taylor & Francis Group (1 study, 7.69%).

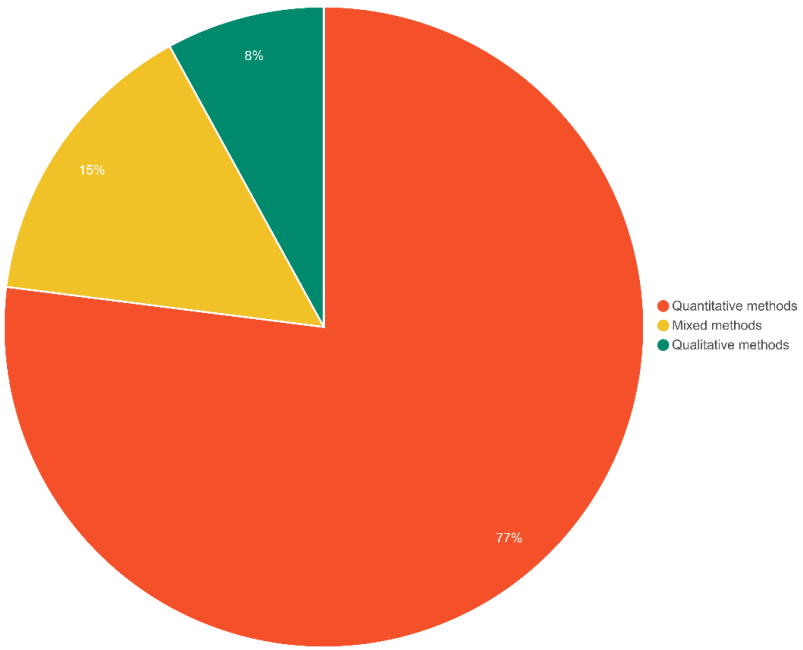
Figure 3.
Publication Sources of Selected Studies



3. Methodological approaches

Regarding research methodologies in Figure 4, the majority of the studies followed a quantitative approach (10 studies,77%), while a smaller portion used mixed methods (2 studies,15%) and qualitative methods (1 study,8%). Quasi-experimental and experimental designs were prevalent among the quantitative studies, often relying on surveys like PLS-SEM modeling and online questionnaires for data collection. Mixed-methods studies typically combined pre- and post-tests with surveys and open-ended questions. Case studies served as the primary means of data gathering for qualitative. Further details on the research methods can be referenced in the appendix.

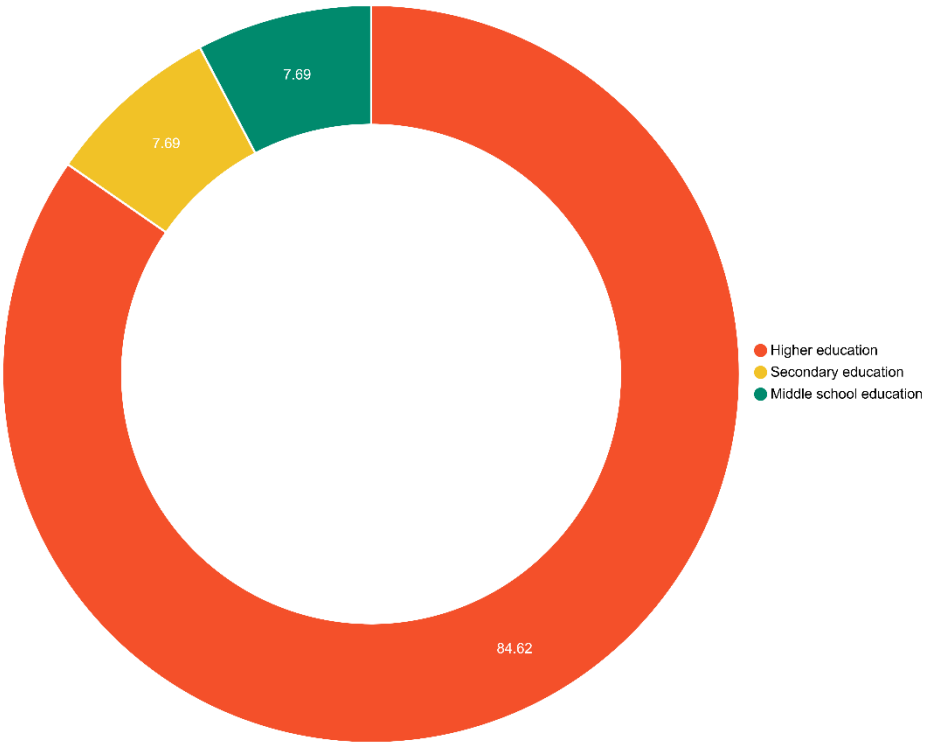
Figure 4.
Methodological Approaches of Selected Studies



4. Educational levels

By reviewing the selected studies, as shown in Figure 5, it is found that higher education was the most commonly examined level, accounting for 84.62% (n = 11). This was followed by secondary education, representing 7.69% (n = 1), and middle school education, also at 7.69% (n = 1). No studies focused on primary, preschool, or kindergarten levels.

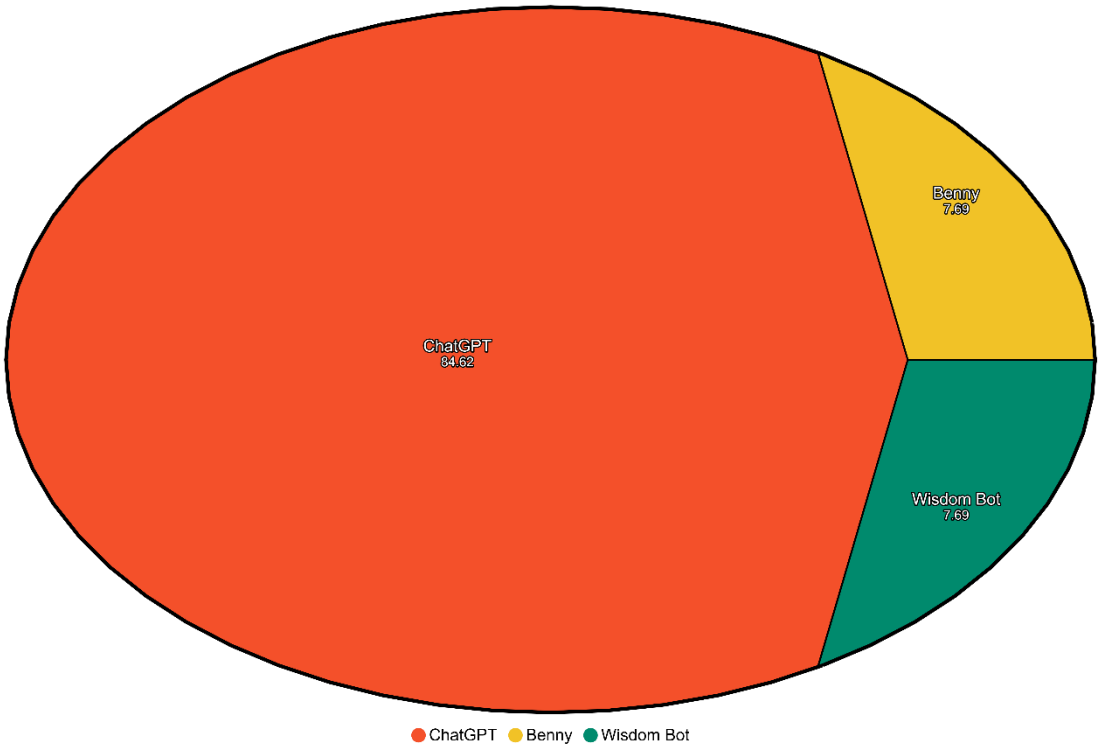
Figure 5.
Educational Levels of Selected Studies



5. Generative AI platforms and apps

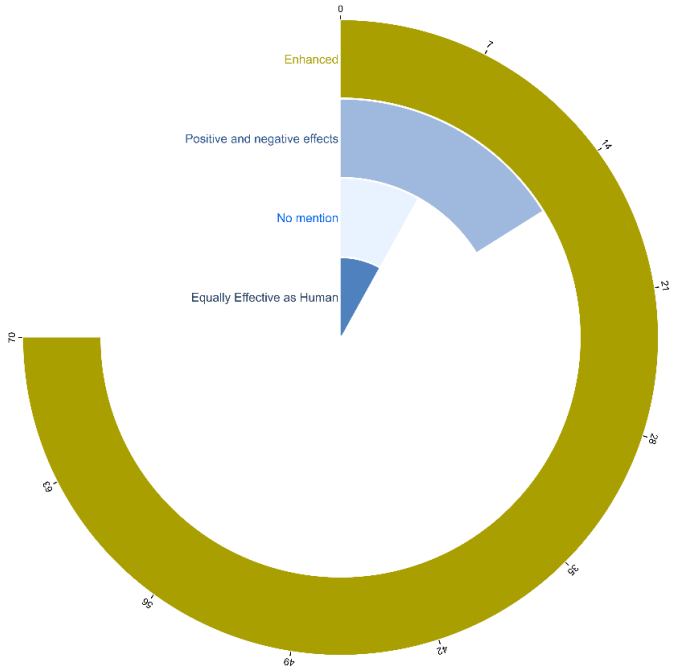
As shown in Figure 6, the most frequently used generative AI platform in the studies was ChatGPT (including GPT-3.5 and GPT-4), accounting for 84.62% (n = 11). Benny and Wisdom Bot were each used in 7.69% (n = 1) of the studies.

Figure 6.
Generative AI Platforms and Apps of Selected Studies



Effective of Generative AI in enhancing educational outcomes (RQ2)
In this section, we reviewed 13 studies. Of these, 9 studies (70%) reported that generative AI effectively enhances learning outcomes. Two studies (15%) found positive and negative effects, suggesting generative AI can boost learning in some areas while hindering it in others. One study (7.5%) concluded that generative AI is as effective as human instruction. Another study (7.5%) did not mention the impact of generative AI on learning outcomes. Notably, none of the studies reported solely negative outcomes from the use of generative AI in education (See Figure 7).

Figure 7.
Effects of Generative AI in Enhancing Education



Factors Influencing the Effect of Generative AI in Education (RQ3)
In addition to educational outcomes, the studies reviewed explored a range of factors impacting the effective of generative AI in education, which were categorized into the following themes:

1. Pedagogical factors, including fairness and ethics, educator involvement, active learning strategies, customization of learning content, instructor guidance, scaffolding, self-regulated learning, personalization based on student knowledge, adaptive learning, and autonomous learning.
2. Psychological factors, such as trust, social presence, motivation, willingness to engage, flow experience, and gamification.
3. Cognitive factors, covering self-efficacy, creativity, contextual relevance, task complexity, cognitive load, real-time feedback, iterative engagement, reflective thinking, and reducing overreliance on AI.
4. Learner factors, including student preferences, interaction frequency, quality of prompts, active engagement, and student acceptance.
5. Technological factors, including personalized interaction, clarity of feedback, prompt engineering, content accuracy, AI hallucinations, digital literacy, access to diverse information, and problem-solving guidance.
6. Institutional factors, such as institutional support, scalability, and efficient resource use.

DISCUSSION AND CONCLUSION

This article investigates how generative AI enhances specific educational outcomes through a systematic review, addressing three key research questions focused on patterns in the reviewed studies. These patterns include study locations, publication sources, research methodologies, educational levels, and generative AI platforms and tools. Additionally, the review assessed the effectiveness of generative AI in improving educational outcomes and examined the factors influencing its implementation in education.

The analysis of study patterns aligns with findings from Guo et al. (2023) and Zainuddin et al. (2020). Most studies on generative AI's impact on learning outcomes were conducted in China. Asian researchers seem more interested in exploring generative AI than in other areas. Furthermore, two Springer journals, *Education and Information Technologies* and *The International Journal of Educational Technology in Higher Education*, frequently publish studies on generative AI's impact, highlighting their authority in the field. Most studies rely on quantitative

approaches, particularly quasi-experimental and experimental designs, although mixed-method approaches are gaining prominence. Contrastingly, qualitative research is less commonly employed. The research also primarily focuses on higher education, with fewer studies conducted at other levels.

Additionally, there is a strong preference for studying ChatGPT, which underscores its powerful capabilities and high user acceptance, while similar software has received less attention. Future research should broaden the scope to include diverse educational contexts (Tafazoli, 2024). More qualitative and mixed-method research is needed to gain deeper insights into how generative AI influences cognitive and emotional learning processes (Wang et al., 2024). Longitudinal studies could also explore the long-term effects of generative AI across different educational levels (Lodge et al., 2023). Finally, a broader investigation of generative AI platforms beyond ChatGPT could provide a more comprehensive understanding of their effectiveness in education (Aydın & Karaarslan, 2023).

Most studies indicate positive impacts for generative AI's effectiveness in enhancing educational outcomes. This indication is likely due to its ability to provide personalized feedback, adapt to individual learning paces, and engage students innovatively. However, studies reporting mixed results suggest that while generative AI excels in delivering quick information or generating content, it may hinder deeper cognitive processes like critical thinking and creativity when over-relied upon. One study found that AI can replicate traditional teaching methods but may not surpass them in all contexts. Notably, the absence of studies reporting only negative outcomes suggests that generative AI adds value to educational settings. Future research should explore its effects on critical thinking and creativity, where mixed results have been observed (Habib et al., 2024; Ruiz-Rojas et al., 2024). Further studies comparing generative AI with human instruction across different educational levels and subject areas would deepen understanding of its complementary or supplementary role (Chan & Tsi, 2024). Lastly, research on the potential drawbacks of over-reliance on AI, particularly in complex cognitive tasks, would help develop balanced and effective educational strategies (Zhai et al., 2024).

Several key factors influence the effectiveness of generative AI in education, consistent with Valverde-Berrocoso et al. (2022). First, pedagogical factors, such as educator involvement, personalized learning, and strategies like scaffolding and active learning, were widely discussed. These elements enable generative AI to meet students' individual needs and support autonomous learning. The customization of content and instructors' role in guiding AI use were also essential for maximizing its potential. Second, psychological factors like trust, motivation, and social presence shape how students engage with AI. When students trust the AI and feel socially connected, they engage more meaningfully. Concepts such as flow experience and gamification further enhance motivation. Third, cognitive factors, including self-efficacy, creativity, and cognitive load, were noted. Generative AI's real-time feedback and iterative engagement enhance cognitive processes, though over-reliance may limit creativity and critical thinking in tasks requiring deep reflection. Fourth, learner factors, such as student preferences, interaction frequency, and the quality of prompts, were crucial for effectiveness. Active engagement and acceptance of AI lead to better learning outcomes. Fifth, technological factors like AI accuracy, clarity of feedback, and prompt engineering were highlighted. Challenges like AI hallucinations and digital literacy were also important, as they can enhance or limit AI's effectiveness. Finally, institutional factors, such as institutional support, scalability, and resource efficiency, were critical for successful AI integration in education. Institutions providing adequate support see better outcomes from AI technology. These factors offer a comprehensive understanding of the dimensions influencing generative AI's effectiveness. Future research should delve into these factors to optimize AI's role in education.

This systematic review has synthesized empirical studies on the effectiveness of generative AI in education, revealing both positive and mixed impacts. While most studies highlight generative AI's capacity to enhance personalized learning, motivation, and engagement, challenges such as over-reliance on AI and reduced creativity in complex tasks persist. Pedagogical, psychological, cognitive, learner, technological, and institutional factors significantly shape AI's effectiveness in educational settings. Future research should explore a broader range of AI tools, methodologies, and educational contexts to deepen understanding and ensure balanced educational system integration.

The main limitation of this review was its reliance on WoS as the sole database for selecting studies. Future reviews could expand to include Scopus and ERIC to broaden the range of sources. Additionally, the small number of articles limited the analysis. Future research should comprise more diverse studies to provide a broader perspective. This review focused solely on generative AI, and future studies may consider exploring other educational technologies or investigating integrating AI with additional tools to enhance educational outcomes.

REFERENCES

Aydın, Ö., & Karaarslan, E. (2023). Is ChatGPT leading generative AI? what is beyond expectations? *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4341500

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

Blagoev, I., Vassileva, G., & Monov, V. (2023). From data to learning: The scientific approach to AI-Enhanced Online Course Design. *2023 International Conference on Big Data, Knowledge and Control Systems Engineering (BdKCSE)*. https://doi.org/10.1109/bdkcse59280.2023.10339693

Chan, C. K., & Tsi, L. H. Y. (2024). Will generative AI replace teachers in Higher Education? A study of teacher and student perceptions. *Studies in Educational Evaluation*, *83*, 101395. https://doi.org/10.1016/j.stueduc.2024.101395

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. *Education and Information Technologies*. https://doi.org/10.1007/s10639-024-12553-x

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

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). https://doi.org/10.1186/s41239-023-00425-2

Fui-Hoon Nah, F., Zheng, R., Cai, J., Siau, K., & Chen, L. (2023). Generative AI and ChatGPT: Applications, challenges, and AI-Human collaboration. *Journal of Information Technology Case and Application Research*, *25*(3), 277–304. https://doi.org/10.1080/15228053.2023.2233814

Guo, K., Zhong, Y., Zainuddin, Z., & Chu, S. K. (2023). Applying game-related methods in the writing classroom: A scoping review. *Education and Information Technologies*, *29*(4), 4481–4504. https://doi.org/10.1007/s10639-023-11998-w

Habib, S., Vogel, T., Anli, X., & Thorne, E. (2024). How does Generative Artificial Intelligence Impact Student creativity? *Journal of Creativity*, *34*(1), 100072. https://doi.org/10.1016/j.yjoc.2023.100072

Hakiki, M., Fadli, R., Samala, A. D., Fricticarani, A., Dayurni, P., Rahmadani, K., & Astiti, A. D. (2023). Exploring the impact of using ChatGPT on student learning outcomes in technology learning: The comprehensive experiment. *Advances in Mobile Learning Educational Research*, *3*(2), 859–872. https://doi.org/10.25082/amler.2023.02.013

Hsu, M.-H. (2023). Mastering medical terminology with ChatGPT and Termbot. *Health Education Journal*, *83*(4), 352–358. https://doi.org/10.1177/00178969231197371

Kadaruddin, K. (2023). Empowering education through generative AI: Innovative Instructional Strategies for tomorrow's learners. *International Journal of Business, Law, and Education*, *4*(2), 618–625. https://doi.org/10.56442/ijble.v4i2.215

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*. https://doi.org/10.35542/osf.io/5er8f

Kurtz, G., Amzalag, M., Shaked, N., Zaguri, Y., Kohen-Vacs, D., Gal, E., Zailer, G., & Barak-Medina, E. (2024). Strategies for integrating generative AI into Higher Education: Navigating challenges and leveraging opportunities. *Education Sciences*, *14*(5), 503. https://doi.org/10.3390/educsci14050503

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

Li, Y., Sadiq, G., Qambar, G., & Zheng, P. (2024). The impact of students' use of ChatGPT on their research skills: The mediating effects of autonomous motivation, engagement, and self-directed learning. *Education and Information Technologies*. https://doi.org/10.1007/s10639-024-12981-9

Lodge, J. M., Thompson, K., & Corrin, L. (2023). Mapping out a research agenda for Generative Artificial Intelligence in tertiary education. *Australasian Journal of Educational Technology*, *39*(1), 1–8. https://doi.org/10.14742/ajet.8695

Martín-Martín, A., Orduna-Malea, E., Thelwall, M., & López-Cózar, E. D. (2018). Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories. *Journal of Informetrics*, 12(4), 1160–1177.https://doi.org/10.31235/osf.io/42nkm

Ming, G. K., & Mansor, M. (2023). Exploring the impact of ChatGPT on teacher professional development: Opportunities, challenges, and implications. *Asian Journal of Research in Education and Social Sciences*. https://doi.org/10.55057/ajress.2023.5.4.6

Montenegro-Rueda, M., Fernández-Cerero, J., Fernández-Batanero, J. M., & López-Meneses, E. (2023). Impact of the implementation of ChatGPT in education: A systematic review. *Computers*, *12*(8), 153. https://doi.org/10.3390/computers12080153

Nguyen Thanh, B., Vo, D. T., Nguyen Nhat, M., Pham, T. T., Thai Trung, H., & Ha Xuan, S. (2023). Race with the machines: Assessing the capability of Generative AI in solving authentic assessments. *Australasian Journal of Educational Technology*, *39*(5), 59–81. https://doi.org/10.14742/ajet.8902

Ogunleye, B., Zakariyyah, K. I., Ajao, O., Olayinka, O., & Sharma, H. (2024). A systematic review of Generative AI for teaching and learning practice. education sciences, 14(6), 636. https://doi.org/10.3390/educsci14060636

Pereira, E., Nsair, S., Pereira, L. R., & Grant, K. (2024). Constructive alignment in a graduate-level project management course: An innovative framework using large language models. *International Journal of Educational Technology in Higher Education*, *21*(1). https://doi.org/10.1186/s41239-024-00457-2

Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems*, *3*, 121–154. https://doi.org/10.1016/j.iotcps.2023.04.003

Ruiz-Rojas, L. I., Salvador-Ullauri, L., & Acosta-Vargas, P. (2024). Collaborative working and critical thinking: Adoption of generative artificial intelligence tools in higher education. *Sustainability*, *16*(13), 5367. https://doi.org/10.3390/su16135367

Sanhita Kar, Chayantika Roy, Meghna Das, Soumita Mullick, & Rupa Saha. (2023). Ai horizons: Unveiling the future of Generative Intelligence. *International Journal of Advanced Research in Science, Communication and Technology*, 387–391. https://doi.org/10.48175/ijarsct-12969

Shahzad, M. F., Xu, S., & Zahid, H. (2024). Exploring the impact of Generative AI-based technologies on learning performance through self-efficacy, fairness & ethics, creativity, and trust in higher education. *Education and Information Technologies*. https://doi.org/10.1007/s10639-024-12949-9

Shi, S. J., Li, J. W., & Zhang, R. (2024). A study on the impact of Generative Artificial Intelligence supported Situational Interactive teaching on students' 'flow' experience and learning effectiveness — a case study of legal education in China. *Asia Pacific Journal of Education*, *44*(1), 112–138. https://doi.org/10.1080/02188791.2024.2305161

Sierra-Correa, P. C., & Cantera Kintz, J. R. (2015). Ecosystem-based adaptation for improving coastal planning for sea-level rise: A systematic review for mangrove coasts. *Marine Policy*, *51*, 385–393. https://doi.org/10.1016/j.marpol.2014.09.013

Tafazoli, D. (2024). Exploring the potential of generative AI in democratizing English language education. *Computers and Education: Artificial Intelligence*, *7*, 100275. https://doi.org/10.1016/j.caeai.2024.100275

Topping, K. J., Douglas, W., Robertson, D., & Ferguson, N. (2022). Effectiveness of online and blended learning from schools: A systematic review. *Review of Education*, *10*(2). https://doi.org/10.1002/rev3.3353

Valverde-Berrocoso, J., Acevedo-Borrega, J., & Cerezo-Pizarro, M. (2022). Educational Technology and student performance: A systematic review. *Frontiers in Education*, *7*. https://doi.org/10.3389/feduc.2022.916502

Verbruggen, S., Depaepe, F., & Torbeyns, J. (2021). Effectiveness of educational technology in early mathematics education: A systematic literature review. *International Journal of Child-Computer Interaction*, *27*, 100220. https://doi.org/10.1016/j.ijcci.2020.100220

Wang, K., Pan, Z., & Lu, Y. (2024). From general AI to custom AI: The effects of generative conversational AI's cognitive and emotional conversational skills on user's guidance. *Kybernetes*. https://doi.org/10.1108/k-04-2024-0894

Wang, W.-S., Lin, C.-J., Lee, H.-Y., Huang, Y.-M., & Wu, T.-T. (2024). Integrating feedback mechanisms and ChatGPT for VR-based experiential learning: Impacts on reflective thinking and Aiot physical hands-on tasks. *Interactive Learning Environments*, 1–18. https://doi.org/10.1080/10494820.2024.2375644

Whittemore, R., & Knafl, K. (2005). The integrative review: Updated methodology. *Journal of Advanced Nursing*, *52*(5), 546-553.

Xie, Z., Wu, X., & Xie, Y. (2024). Can interaction with generative artificial intelligence enhance learning autonomy? A longitudinal study from comparative perspectives of virtual companionship and knowledge acquisition preferences. *Journal of Computer Assisted Learning*, *40*(5), 2369–2384. https://doi.org/10.1111/jcal.13032

Zainuddin, Z., Chu, S. K., Shujahat, M., & Perera, C. J. (2020). The impact of gamification on learning and instruction: A systematic review of empirical evidence. *Educational Research Review*, *30*, 100326. https://doi.org/10.1016/j.edurev.2020.100326

Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. *Smart Learning Environments*, *11*(1). https://doi.org/10.1186/s40561-024-00316-7

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

Zhu, Y., Zhu, C., Wu, T., Wang, S., Zhou, Y., Chen, J., Wu, F., & Li, Y. (2024). Impact of assignment completion assisted by large language model-based chatbot on middle school students' learning. *Education and Information Technologies*. https://doi.org/10.1007/s10639-024-12898-3

APPENDIX

Article	Countries involved	Publication sources	Research aims	Methodologies	Educational levels	GAI platforms	Effectiveness of GAI in enhancing educational outcomes	Factors influencing GAI effectiveness in education	Themes
1. Shahzad et al., (2024)	China	Education and Information Technologies (Springer)	The study examines how generative AI tools, such as ChatGPT, impact students' learning performance, with a focus on self-efficacy, fairness, creativity, and trust	Quantitative research (surveys and PLS-SEM modeling)	Higher education	ChatGPT and other LLM-based chatbots	Generative AI enhances learning outcomes	Self-efficacy; Creativity; Fairness and ethics; Trust; Personalized interaction	Cognitive factors; Pedagogical factors; Psychological factors; Technological factors;
2. Escalante et al., (2023)	Asia Pacific region	International Journal of Educational Technology in Higher Education (Springer)	The study investigates the efficacy of AI-generated feedback in improving writing skills and ENL students' preferences for AI or	Mixed-methods approach: A quasi-experimental design with pre- and post-tests (Study 1) and surveys and open-ended questions (Study 2)	Higher education	GPT-4	Equally effective compared with human	Clarity and specificity of feedback; Student preference and perception; Prompt engineering; Institutional support and educator involvement	Technological factors; Learner factors; Pedagogical factors;

			human tutor feedback						
3. Xie et al., (2024)	China	Journal of Computer Assisted Learning (Wiley)	To explore the impact of generative AI interaction on learning autonomy and social presence	Quantitative research (online surveys)	Higher education	Benny	Enhancing learning outcomes as a virtual companion; Reducing learning outcomes as a knowledge acquisition tool	Interaction frequency with AI; Social presence; Learner preferences; Technological capabilities	Learner factors; Psychological factors; Technological factors;
4.Pereira et al., (2024)	Canada	International Journal of Educational Technology in Higher Education (Springer)	Using LLMs, particularly ChatGPT, to develop a constructive alignment framework for course design.	Qualitative research (case study)	Higher education	ChatGPT	Generative AI (GAI) enhancing learning outcomes	Quality of prompts; Adaptability to course structures; Instructor oversight; Active learning strategy; Contextual relevance and specificity;	Technological factors; Pedagogical factors; Cognitive factors;
5. Zhu et al., (2024)	China	Education and Information Technologies (Springer)	Investigating the impact of LLM-based chatbots on middle school student performance	Quantitative research (quasi-experiments)	Middle school	Wisdom Bot	Beneficial for task completion but potentially harmful for deep learning	Quality of prompts and student interaction; Instructor guidance and scaffoldi	Learner factors; Pedagogical factors; Technological factors; Cognitive factors; Psychol

							ng; AI's content accuracy and hallucinations; Task complexity and cognitive load; students' motivation and willingness to engage with AI	ogical factors;
6. Shi et al., (2024)	China	Asia Pacific Journal of Education (Taylor & Francis Group)	Investigates the impact of Generative AI-supported Situational Interactive Teaching on students' flow experiences and learning effectiveness	Quantitative approach (quasi-experimental)	Higher education	GPT-3.5 large-scale language model	GAI-supported situational interactive teaching significantly enhances students' learning outcomes	Integration of contextual and interactive learning; Enhancement of flow experience; Active student engagement with AI tools; Scalability and efficient resource use	Pedagogical factors; Psychological factors; Learner factors; Institutional factors:
7. Hsu, (2023)	Taiwan, China	Health Education Journal (Sage)	To assess the effectiveness of ChatGPT in improving medical terminology understa	Quantitative research (experimental research)	Higher education	ChatGPT	ChatGPT is found to be effective in enhancing learning outcomes	Personalization of learning based on student knowledge level;Time dedicated to interaction with	Pedagogical factors; Cognitive factors; Psychological factors; Technological factors;

			nding				AI; Engagement through gamification and ease of use; Clarity and depth of explanations provided by AI	
8. Dalgıç et al., (2024)	Turkey	Journal of Hospitality, Leisure, Sport & Tourism Education (Elsevier)	To investigate how ChatGPT, digital literacy, and individualized learning in tourism education influence learning outcomes	Quantitative research (experimental research)	Higher education	ChatGPT	ChatGPT is effective in enhancing learning outcomes	Digital literacy; Engagement and motivation; Real-time feedback and instant access to information; Access to culturally diverse information	Technological factors; Psychological factors; Cognitive factors; Institutional factors;
9. Zhou & Kim., (2024)	South Korea	Education and Information Technologies (Springer)	To assess how ChatGPT-4 influences student learning in music education	Quantitative research (quasi-experimental design)	Higher education	ChatGPT-4	ChatGPT-4 effective in improving educational outcomes in music education	Instant feedback and quick response; Customization of learning content; Engagement through interactive features; Access to a wide range of music knowledge;	Cognitive factors; Pedagogical factors; Psychological factors; Technological factors; Institutional factors;

								Student acceptance and comfort with AI	
10. Chen & Chang., (2024)	Taiwan, China	Education and Information Technologies (Springer)	To examines the effectiveness of AI-assisted game-based learning on students' science learning outcomes	Quantitative research (quasi-experimental design)	Secondary education	ChatGPT	AI-assisted game-based learning significantly enhances educational outcomes	Real-time feedback for immediate learning adjustments and reducing cognitive load for better learning; Adaptive learning and personalized responses; Interactive engagement enhances effectiveness; Problem-solving guidance enhances AI functionality	Cognitive factors; Pedagogical factors; Psychological factors; Technological factors;
11. Wang et al., (2024)	Taiwan, China	Interactive Learning Environments (Taylor & Francis Group)	To examine the role of ChatGPT in providing feedback in VR-based experiential learning	Mixed research (randomized controlled trial (RCT) and hands-on tasks, and reflective thinking)	Higher education,	ChatGPT	No mention educational outcomes	Real-time personalized feedback; Prompting reflective thinking; Reducing cognitive overload;	Cognitive factors; Pedagogical factors; Technological factors;

								Adaptability to learner queries; Simulating teacher and peer interaction	
12. Lee et al., (2024)	Taiwan, China	International Journal of Educational Technology in Higher Education (Springer)	To investigates the effect of Guidance-based ChatGPT-Assisted Learning Aid (GCLA) with a guidance mechanism in a blended learning environment	Quantitative research (randomized controlled Trial (RCT)	Higher education	ChatGPT	The GCLA improve learning outcomes through providing guidance rather than answers by ChatGPT	Guidance mechanism; Reducing overreliance on AI; Supporting self-regulated learning	Cognitive factors : Pedagogical factors;
13. Li et al., (2024)	Pakistan	Education and Information Technologies (Springer)	To examines the impact of ChatGPT usage on students' research skills	Quantitative study (quasi-experimental design)	Higher education	ChatGPT	Significant positive effects of ChatGPT on these educational outcomes	Personalized feedback; Specific prompts; Autonomous learning; Iterative engagement; Reducing cognitive load	Technological factors; Pedagogical factors; Cognitive factors;