International Society of the Learning Sciences

ChatGPT as a Feedback Provider in Digital Learning Games: A Systematic Literature Review

Mahboobeh Mehrvarz, Bruce M. McLaren, Autumn Qiu
mmehrvar@andrew.cmu.edu, bmclaren@andrew.cmu.edu, autumnq@andrew.cmu.edu
Carnegie Mellon University
Krystof Purtell, University of Pennsylvania, krystofp@sas.upenn.edu
Sila Yesilyurt, Cornell University, syesilyu@andrew.cmu.edu
Doan Viet Nguyen, Cornell University, dtn26@cornell.edu

Abstract: This systematic review explores how ChatGPT has been integrated into digital learning games, focusing on its role as a feedback provider. Analyzing 17 empirical studies, the review found that ChatGPT is mainly used to deliver cognitive and motivational feedback, often leading to improved engagement and learning. While students generally appreciate the immediate support, concerns remain about the accuracy and depth of its responses.

Introduction

The integration of Large Language Models (LLMs) like ChatGPT into educational technology has received growing attention (Yan et al., 2024). LLMs offer several benefits, including question generation, feedback provision, and essay grading (Nguyen et al., 2023). ChatGPT's ability to deliver adaptive, personalized feedback positions it to transform digital learning, especially within interactive learning games (Lanzi & Loiacono, 2023).

Vygotsky's Zone of Proximal Development (ZPD) highlights the gap between what learners can do alone and with guidance (Vygotsky, 1978). Within this framework, ChatGPT can act as a scaffold, helping students overcome challenges and reach new learning goals. Its real-time analysis of learner input enables adaptive feedback that promotes knowledge-building and problem-solving (Gill & Kaur, 2023; Goslen et al., 2024).

ChatGPT delivers cognitive, metacognitive, and motivational feedback—improving understanding, encouraging reflection, and supporting engagement (McLaren et al., 2022; Hassan et al., 2019). It can also explain concepts, answer questions, and adjust challenge levels based on student performance (Chen & Hou, 2024). Despite these benefits, ChatGPT may provide inaccurate or shallow feedback, leading to frustration. Ethical concerns and limited context-awareness also pose challenges (Chen & Chang, 2024). These issues call for further research to improve its adaptability and effectiveness.

Given ChatGPT's strengths and limitations, understanding students' perspectives is key to evaluating its effectiveness (Nazari & Saadi, 2024). How students experience its cognitive and motivational feedback—whether helpful or hindering—is critical to shaping its educational role. Despite some existing studies, no systematic review has explored ChatGPT as a feedback provider in digital learning games. This review addresses that gap by examining its impact on student engagement and learning through three key research questions:

RQ1: What specific types of adaptive feedback does ChatGPT provide in digital games, particularly focusing on cognitive and motivational feedback?

RQ2: What is the overall effectiveness of ChatGPT as a feedback provider in digital learning games?

RQ3: What are students' opinions and attitudes regarding ChatGPT as a feedback provider in digital learning games?

Methods and materials

This systematic review followed a structured, multi-phase approach—identification, screening, and eligibility—supplemented by quality assessment and snowballing to ensure comprehensive coverage.

Search strategy

Initial exploratory searches informed the search terms, focused on ChatGPT and digital learning games to align with the research questions. ChatGPT was the sole AI tool included due to its wide use, accessibility, and consistent interface. Since its 2022 release, it has been widely studied, providing a strong foundation for review.

We included empirical studies using ChatGPT in digital learning games across subjects like math, science, and programming. Game types included simulations, quizzes, and role-playing, with learners from middle school to adults.

We used Boolean search strings (e.g., "ChatGPT" AND "digital games" AND "learning"), adjusted by the database. Searches were conducted across Scopus, Web of Science, Wiley, IEEE, ACM, PsycINFO, and ScienceDirect for broad interdisciplinary coverage.

Study review process

The initial search returned 1,852 records. After filtering by language and date (2018–2024) and removing duplicates, 1,318 studies remained.

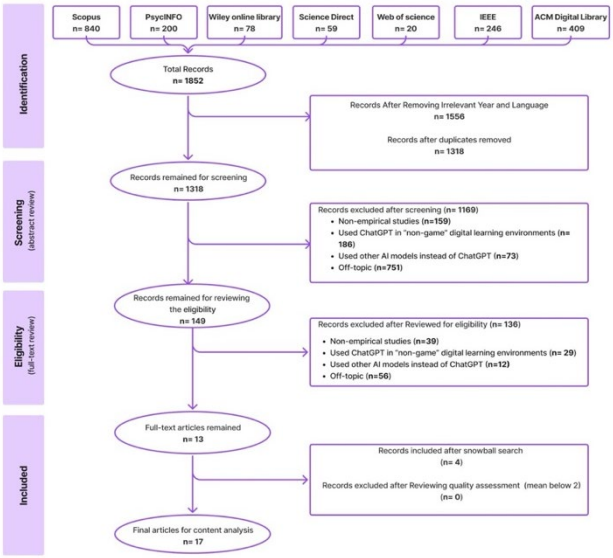
During screening, abstracts were reviewed, and full texts were examined as needed. In the eligibility phase, two authors independently reviewed the remaining full texts and abstracts. Studies that did not meet the inclusion criteria were excluded during both phases. Ultimately, 17 studies were retained for the systematic review (see Figure 1). To ensure comprehensive coverage, we applied backward snowballing (examining reference lists) and forward snowballing (citation tracking) during the final stage.

Quality assessment

The quality of the studies was assessed using criteria from Theelen et al. (2019). Each study was rated on a scale from 0 to 3 based on research clarity, methodological rigor, data collection, and relevance. A second researcher independently reviewed the studies, and average scores were calculated. All 17 studies met the minimum threshold (score ≥ 2) and were included. Relevant data were then extracted to address the research questions.

Figure 1

PRISMA Flow Diagram



Results

Table 1 summarizes some included studies, highlighting subject matter, feedback format, integration mode, feedback type, and performance related to ChatGPT.

Table 1

A Sample of Selected Studies in this Systematic Review (N = 17)

Authors & Year	Subject Matter	Feedback Format	Integration Mode	Feedback Type	Performance of the feedback provided
Chen & Chang (2024)	Science	Text-only	External interface	Cognitive & Motivational	Positive
Chien et al. (2024)	Science	Text-only	External interface	Cognitive & Motivational	Positive
Nguyen et al. (2023)	Math	Text-only	Not integrated into gameplay	Cognitive, Metacognitive & Motivational	Positive & Negative
French et al. (2023)	Programming	Text-only	Partially embedded	Cognitive	Positive & Negative

Across studies, feedback was consistently text-based, with no multimodal outputs reported, except for one (Pears et al., 2023), which converted responses to audio using text-to-speech. Feedback integration varied: in some cases (e.g., French et al., 2023), ChatGPT was embedded in the game interface; in others (e.g., Chen &

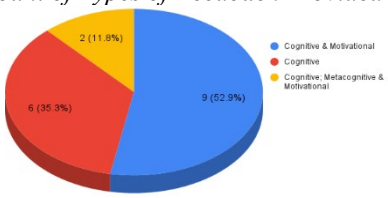
Chang, 2024), it was an external interface accessed via a separate chat window. Some studies, like Nguyen et al., 2023, used ChatGPT offline to analyze feedback potential post hoc.

RQ1: What specific types of adaptive feedback does ChatGPT provide in digital games, particularly focusing on cognitive and motivational feedback?

As shown in Figure 2, most feedback fell into cognitive and motivational categories (52.9%), highlighting ChatGPT's support for both understanding and emotional engagement. Cognitive feedback alone accounted for 35.3%, while only 11.8% involved a mix of cognitive, metacognitive, and motivational support, indicating that comprehensive feedback is less commonly implemented.

Figure 2

Count of Types of Feedback Provided



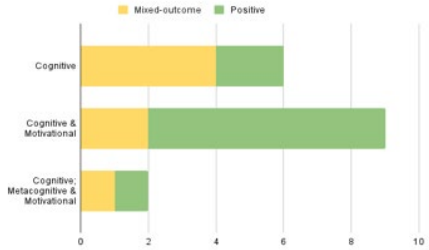
RQ2: What is the overall effectiveness of ChatGPT as a feedback provider in digital learning games?

Figure 3 shows outcomes across three adaptive feedback types: cognitive, cognitive and motivational, and cognitive, metacognitive, and motivational. ChatGPT's effectiveness was classified as positive (showing clear learning gains) or mixed (combining positive, neutral, or negative effects).

The cognitive and motivational category had the strongest results, with 7 positive and 2 mixed outcomes, suggesting overall effectiveness. In contrast, cognitive feedback showed mixed results (4 mixed, 2 positive), indicating varied impact depending on context. The comprehensive feedback category appeared only twice (1 positive, 1 mixed), reflecting limited use and unclear impact in the current dataset.

Figure 3

Distribution of Outcomes by Adaptive Feedback Types



RQ3: What are students' opinions and attitudes regarding ChatGPT as a feedback provider in digital learning games?

Table 2

A Sample of Students' Perceptions of ChatGPT

Authors	Overall Perception	Data Collection Method	Reasons
Chen & Chang (2024)	Positive	Interview	- Enhanced Learning and Understanding - Performing well as an evaluator
Mezzaro et al., (2024)	Positive & Negative	Survey	- Helpful for generic questions - Poor for code examples - Fails to meet students' high expectations
Chen & Hou (2024)	Positive, Neutral & Negative	Scales	-Enhancing engagement and learning -Some students struggled with formulating questions - Lack of helpfulness

The final research question explored students' opinions of ChatGPT as a feedback provider in digital learning games. Perceptions were gathered through interviews, surveys, and questionnaires. For example, Chen and Chang (2024) used focus groups with open-ended questions to understand students' experiences with ChatGPT and the game.

As shown in Table 2, perceptions ranged from positive to neutral and negative. Many students (e.g., Chien et al., 2024) reported that ChatGPT enhanced their learning, especially when it provided accurate evaluations. Others (e.g., Mezzaro et al., 2024; Stampfl et al., 2024) had mixed views—appreciating ChatGPT's interactivity but criticizing its limited depth, weak code examples, and lack of emotional connection.

Neutral responses (e.g., Jost, 2024) reflected limited usefulness, often due to low emotional engagement or difficulty in forming questions. Some studies (e.g., French et al., 2023) raised concerns about ChatGPT's computational accuracy, especially in math-related tasks.

Discussion and conclusion

While several studies have explored ChatGPT as a support tool in digital learning games, no prior systematic review has examined this topic comprehensively. This review addresses that gap by analyzing the types of feedback ChatGPT provides, its impact on learning, and students' perceptions.

Most studies focused on cognitive and motivational feedback, with a few also addressing metacognitive aspects (e.g., Ho & Lee, 2023). Overall, the findings were positive, suggesting ChatGPT can enhance engagement, motivation, and learning behaviors. For instance, Chen and Chang (2024) found that ChatGPT-supported games improved motivation and reduced cognitive load in science learning. However, some studies (e.g., Chen & Hou, 2024) noted that ChatGPT did not outperform traditional methods, indicating its strength may lie more in engagement than in direct learning gains.

Studies focusing solely on cognitive feedback showed mixed results. Some reported improved understanding (e.g., Nikolaeva et al., 2023), while others (e.g., Mezzaro et al., 2024) highlighted risks such as overreliance and unrealistic expectations. These findings suggest the need for refining ChatGPT's feedback to better suit diverse learner needs and task complexities.

A smaller set of studies (e.g., Nguyen et al., 2023; Liang et al., 2024) explored comprehensive feedback combining cognitive, metacognitive, and motivational support. While Liang et al. (2024) reported improvements in metacognition and engagement, Nguyen et al. (2023) found inconsistencies—ChatGPT performed well with conceptual questions but struggled with complex tasks. Feedback lacked consistency in motivational phrasing, suggesting further research is needed to evaluate ChatGPT's ability to deliver holistic support effectively.

Regarding the final research question—students' opinions of ChatGPT as a feedback provider—findings revealed varied perceptions, from positive to neutral and negative, depending on its role in the learning experience. Many studies reported positive views, highlighting ChatGPT's effectiveness in supporting understanding, creativity, and technical skills like programming (e.g., Chen & Chang, 2024; French et al., 2023). These strengths position ChatGPT as a promising tool for scaffolding and engagement, aligning with Vygotsky's Zone of Proximal Development (ZPD), which emphasizes guided support for learning.

However, several studies noted limitations. In complex tasks such as code generation, students expressed frustration with ChatGPT's accuracy and depth (e.g., Mezzaro et al., 2024). Others pointed to its lack of emotional engagement and negotiation (Stampfl et al., 2024). Some even preferred human feedback over LLMs (Nazaretsky et al., 2024), raising questions about whether ChatGPT can fulfill the adaptive role envisioned in ZPD.

Overall, ChatGPT's effectiveness is context-dependent—strong in structured cognitive tasks but limited in areas needing social or emotional sensitivity. While useful in digital learning games, it cannot replace human interaction, especially for complex feedback. Human oversight remains essential. Future research should enhance ChatGPT's support for higher-order thinking and explore integrated models of human-AI feedback.

References

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*, 1-22.

Chen, Y. C., & Hou, H. T. (2024). A Mobile Contextualized Educational Game Framework With ChatGPT Interactive Scaffolding for Employee Ethics Training. *Journal of Educational Computing Research*, *62*(7), 1737-1762.

Chien, C. C., Chan, H. Y., & Hou, H. T. (2024). Learning by playing with generative AI: design and evaluation of a role-playing educational game with generative AI as scaffolding for instant feedback interaction. *Journal of Research on Technology in Education*, 1-20.

French, F., Levi, D., Maczo, C., Simonaityte, A., Triantafyllidis, S., & Varda, G. (2023). Creative use of OpenAI in education: case studies from game development. *Multimodal Technologies and Interaction*, *7*(8), 81.

Gill, S. S., & Kaur, R. (2023). ChatGPT: Vision and challenges. *Internet of Things and Cyber-Physical Systems*, *3*, 262-271.

Goslen, A., Kim, Y. J., Rowe, J., & Lester, J. (2024). Llm-based student plan generation for adaptive scaffolding in game-based learning environments. *International Journal of Artificial Intelligence in Education*, 1-26.

Hassan, L., Dias, A., & Hamari, J. (2019). How motivational feedback increases user's benefits and continued use: A study on gamification, quantified-self and social networking. *International Journal of Information Management*, *46*, 151-162.

Ho, W., & Lee, D. (2023). Enhancing Engineering Education in the Roblox Metaverse: Utilizing chatGPT for Game Development for Electrical Machine Course. *International Journal on Advanced Science, Engineering & Information Technology*, *13*(3).

Jost, P. (2024). AI-deation: When the Teacher is a Transformer in Role-Playing to create Privacy Decision Serious Games. International Journal of Serious Games, 11(2), 65-86.

Lanzi, P. L., & Loiacono, D. (2023, July). Chatgpt and other large language models as evolutionary engines for online interactive collaborative game design. In *Proceedings of the Genetic and Evolutionary Computation Conference* (pp. 1383-1390).

Liang, H. Y., Hwang, G. J., Hsu, T. Y., & Yeh, J. Y. (2024). Effect of an AI-based chatbot on students' learning performance in alternate reality game-based museum learning. *British Journal of Educational Technology*.

Mezzaro, S., Gambi, A., & Fraser, G. (2024, June). An empirical study on how large language models impact software testing learning. In *Proceedings of the 28th International Conference on Evaluation and Assessment in Software Engineering* (pp. 555-564).

McLaren, B. M., Richey, J. E., Nguyen, H., & Hou, X. (2022). How instructional context can impact learning with educational technology: Lessons from a study with a digital learning game. *Computers & Education*, *178*, 104366.

Nazaretsky, T., Mejia-Domenzain, P., Swamy, V., Frej, J., & Käser, T. (2024, September). AI or human? Evaluating student feedback perceptions in higher education. In *European Conference on Technology Enhanced Learning* (pp. 284-298). Cham: Springer Nature Switzerland.

Nazari, M., & Saadi, G. (2024). Developing effective prompts to improve communication with ChatGPT: a formula for higher education stakeholders. *Discover Education*, *3*(1), 45.

Nguyen, H. A., Stec, H., Hou, X., Di, S., & McLaren, B. M. (2023, August). Evaluating chatgpt's decimal skills and feedback generation in a digital learning game. In European Conference on Technology Enhanced Learning (pp. 278-293). Cham: Springer Nature Switzerland.

Nikolaeva, S., Chernysh, V., Boiko, H., & Galynska, O. (2023). BUSINESS SIMULATION GAMES IN DEVELOPING PRE-SERVICE RESTAURANT MANAGERS'SPEECH INTERACTION COMPETENCE USING GPTChat. *Advanced Education*, 27-37.

Pears, M., Poussa, C., & Konstantinidis, S. T. (2023). Progressive Healthcare Pedagogy: An Application Merging ChatGPT and AI-Video Technologies for Gamified and Cost-Effective Scenario-Based Learning. In *Interactive Mobile Communication, Technologies and Learning* (pp. 106-113). Cham: Springer Nature Switzerland.

Stampfl, R., Geyer, B., Deissl-O'Meara, M., & Ivkić, I. (2024). Revolutionising Role-Playing Games with ChatGPT. *arXiv preprint arXiv:2407.02048*.

Theelen, H., Van den Beemt, A., & den Brok, P. (2019). Classroom simulations in teacher education to support preservice teachers' interpersonal competence: A systematic literature review. *Computers & Education*, *129*, 14-26.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (Vol. 86). Harvard university press.

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

Acknowledgments

We would like to acknowledge the assistance of ChatGPT-4 for providing language editing and proofreading support. The content and ideas presented in this paper are entirely our own.