OPEN PRAXIS

# Bridging the Gap: ChatGPT's Role in Enhancing STEM Education

RESEARCH ARTICLE

LINGFEI LUAN
XI LIN
YAN DAI

*Author affiliations can be found in the back matter of this article

INTERNATIONAL COUNCIL FOR OPEN AND DISTANCE EDUCATION

## ABSTRACT

This paper investigates how ChatGPT, an AI chatbot developed by OpenAI, can be introduced to STEM education, specifically, an "Introduction to Cognitive Neuroscience" class. Using mixed-method research, the study conducted an experiment to collect students' performance scores and their feedback to examine the potential impacts from ChatGPT on their critical thinking skills, long-term retention of knowledge, and group learning interactions through collaborative projects. The results demonstrate significant disparities between AI-generated (ChatGPT) and human input. Upon analyzing the grade fluctuations before and after receiving input, it was shown that students who received feedback from ChatGPT encountered a more significant decrease (median = −12) compared to those who received feedback from humans (median = −5). This result indicates potential shortcomings in the effectiveness of AI feedback. Human feedback has significantly higher Retention Proxy scores than that of ChatGPT feedback, suggesting that human feedback can be potentially more effective in fostering long-term retention of course material. Investigation into collaboration dynamics of learning found that feedback given by humans tends to be more positive (63 vs. 40 in sentiment score) and also more focused on improvement and understanding. The theme of feedback is different between two conditions. That is, human feedback emphasizes scientific and detailed approaches while ChatGPT feedback emphasizes educational aspects and cognitive functions. These results indicate that while ChatGPT has potential benefits on educational settings, human feedback is superior in many ways. This study contributes to the continuing discourse on the use of AI in STEM education and emphasizes the significance of maintaining a harmonious combination of AI and human involvement in delivering educational feedback.

**CORRESPONDING AUTHOR:**
**Lingfei Luan**
Psychological Science, Belmont University, USA
lluan@bgsu.edu

**KEYWORDS:**
ChatGPT; STEM Education; Artificial Intelligence; Critical Thinking; Knowledge Retention; Collaborative Learning; Educational Feedback; Higher Education

**TO CITE THIS ARTICLE:**
Luan, L., Lin, X., & Dai, Y. (2025). Bridging the Gap: ChatGPT's Role in Enhancing STEM Education. *Open Praxis*, *17*(1), pp. 108–128. DOI: https://doi.org/10.55982/openpraxis.17.1.685

## INTRODUCTION

While STEM education remains crucial for students' future success, integrating Artificial Intelligence (AI) in educational environments offers unprecedented opportunities to revolutionize teaching and learning approaches. AI-powered tools can solve some academic problems of student engagement and personalization of the learning process (Bhutoria, 2022; Gligorea et al., 2023; Kaswan et al., 2024; Lin et al., 2024). Particularly in STEM fields, the educational scene gradually suffers from adapting to new technology and labor requirements. Research shows that the workplace and technology are changing at a speed that calls both people with specialized knowledge in the older technologies and also the capacity to adapt to the new ones (Acemoglu, 2002; Autor, 2015; Brynjolfsson & McAfee, 2011). STEM-related jobs are projected to impact the economy significantly more than other occupations (Oleson et al., 2014). This fact underscores the significance of a STEM education in preserving the nation's economic edge and driving invention and creativity (STEM Education for the Future, 2020).

Not only is it necessary to impart subject content, but STEM education is also needed to build critical thinking and problem-solving skills (Vahidy, 2019). These studies are essential for handling the modern world's paradoxes and challenges and simultaneously driving forward discoveries. As computer technology undergoes rapid changes, the STEM curriculum may be the hardest to update satisfactorily (Felder & Brent, 2024). As sectors such as AI, biotechnology, and renewable energy adapt and modernize, educational institutions like schools and colleges modify curricula to keep up with these advancements. This helps address long-standing issues in education. (Liyanage et al., 2022; Rane, 2023; Tekin & van der Schaar, 2015). One can give an example of AI where students' learning results may be adapted to personalized needs and more comfortable with complicated subjects. Some AI applications used in education are making significant advancements. An adaptive learning platform uses a sophisticated algorithm for catering to individual students; AI tutors are more assistive at such stages where students need clarifications and feedback (Ahmad et al., 2020; Maghsudi et al., 2021).

ChatGPT, with its state-of-the-art Natural Language Processing (NLP) capabilities, can effectively communicate, generate content, and offer immediate personalized feedback to its users. It can be used independently or with other STEM teaching materials and domains by enhancing the learning experiences (Grassini, 2023; Gill et al., 2024; Rasul et al., 2023). The application of ChatGPT has led to an innovative technology revolution in feedback distribution to education (Luan et al., 2023). ChatGPT is among one of the highly proficient natural language generation platforms when it comes to human language with higher level logical reasoning and a very impressive level of precision. In addition, it can offer generalization in response to subjection from varying intellectual capacities and the use of pre-existing knowledge. For example, ChatGPT may be helpful as a tech support assistant, coursework, and college lecture help, while it can give individual answers as needed. The invention of such artificial communication systems and the utility of AI tech will show the effect found at that precise point of time.

AI assistants should be a very inclusive tool; they can explain the steps taken in online courses and provide individual details to individual students within a few seconds. It remains a challenging and complex task to develop more effective AI communication tools, and the effect of AI technology will be the main criterion for judging the level of their success. This will warrant a data collection procedure and an efficient judgment system that would rate the performance of human lecturers so that time is not wasted. The practicality of building more effective artificial communication systems and the significance of AI technology will determine the precise result. With the assistance of the powerful storing and collecting abilities, AI-embedding systems can collect data and assess the efficiency of human teachers in order to minimize time waste. Moreover, the model can analyze the correctness and quality of the answers, which would, in turn, ensure that each student is fairly assessed with accurate responses. According to most recent studies, educators who depend on ChatGPT and other digital technologies can significantly speed up grading and feedback (Adıgüzel et al., 2023). On the other hand, ChatGPT can improve the involvement and interaction of the learning process by offering prompt and accurate note assistance to learners (Luan et al., 2023). Meanwhile, Zhong (2023) indicated that a "bridge" between the gap in the present educational method, where the teacher brings course learning materials and gives personalized feedback needed by the learner in the present. An important aspect of the evaluation of the efficiency of the use of AI in instruction may be ethical issues.

Ethical problems could be crucial in assessing how effectively artificial intelligence is used in education. Concerns regarding the ethical aspect of artificial intelligence use in education were voiced by the researchers, who underlined the need to pay especially special attention to issues of accountability, transparency, openness, diversity, equity, privacy, and system security and safety (Ghandour et al., 2024; Waghmare, 2023).These ethical enforcements would ensure the application of AI technology to fulfill ethical and educational purposes, and their significance cannot be exaggerated. While these ethical enforcements would guarantee the implementation of AI to achieve ethical and educational objectives, their importance cannot be overstated. Moreover, integrating AI into the academic system raises concerns over representation, inaccurate attribution, and academic honesty. One study revealed that a significant obstacle in information sharing between humans and computers is the machines' incapability to recognize and attribute actions to specific individuals, resulting in difficulties in generating desired outcomes (Bozkurt et al., 2023). On the other hand, prompt engineering is also equipped with the potential to promote the efficiency of AI in educational feedback (Bozkurt et al., 2023). Properly engineered and skillfully created prompts would assist in producing AI-generated answers of higher quality and, in the proper channels, developing a more engaging and valuable learning process for the students. By applying these strategies, a change in emphasis occurs, resulting in enhanced efficiency in the education sector. Such change also gives rise to ethical considerations. Although it is essential to address these concerns, ensuring that the academic level is not compromised is also crucial. By working on these issues and improving how AI is implemented, it would be possible to bring a perspective of responsibility and efficiency to an AI-aided education.

Therefore, this study seeks to evaluate the extent to which guidance given with ChatGPT systematically enhances critical thinking, reinforces long-term knowledge retention, and promotes collaborative relationships among higher-education students. Through the research questions posed, the authors intend to identify the uncertainties around how AI can contribute to smoothening educational relations and, hence, how this can be used to maximize its proficiency in a particular curriculum design.

## RESEARCH FOCUS AND QUESTIONS

This study aims to examine the effects of ChatGPT feedback on improving critical thinking, long-term information retention, and collaborative learning dynamics in higher education. The research questions posed to guide this inquiry are as follows:

> RQ1: How does ChatGPT feedback influence the development of critical thinking and problem-solving skills in students compared to traditional human feedback? This question addresses the potential of ChatGPT to foster analytical thinking and the application of knowledge in problem-solving scenarios, which are crucial skills in STEM education.
>
> RQ2: What is the effect of ChatGPT feedback on the long-term retention of course material? Knowledge retention is a significant challenge in education, and this question explores whether the immediacy and personalization of ChatGPT feedback can lead to more durable learning outcomes.
>
> RQ3: How does the integration of ChatGPT into group projects affect collaborative learning dynamics and outcomes? Given the importance of teamwork and communication in higher education, this question examines the role of ChatGPT in facilitating and enhancing group interactions and learning processes.

The relevance of each of these questions is situated in the current educational dilemmas and the hypothetical promise of ChatGPT. The present study intends to respond to these questions to contribute to our current understanding regarding the role of AI in education and its pedagogical potential. The discoveries of this research would have significant implications for how ChatGPT can be used as an instructional tool, possible cultural and hybrid learning approaches, and student involvement in higher education. Knowing how ChatGPT influences learning upsides is key for educators, policymakers, and curriculum creators to use this new type of modern technology.

# LITERATURE REVIEW

## THE IMPORTANCE AND EFFECTIVENESS OF FEEDBACK IN LEARNING

The feedback that students receive in educational settings profoundly impacts on their learning experiences and the caliber of their course achievement. These impacts are greatly influenced by the source and manner in which the feedback is provided. The theoretical framework of feedback prioritizes precision and clarity to ensure that students comprehensively understand their present performance status and the necessary steps to enhance their performance. Such precise, unambiguous, and meaningful feedback was highly beneficial in guiding students' learning and improving their performance in reaching learning objectives (Hattie & Timperley, 2007; Luan, 2016; Nicol & Macfarlane-Dick, 2006). Studies that have investigated the impact of feedback supplied by humans in contrast to input provided by an AI tool like ChatGPT have uncovered a variety of findings around student involvement and learning. Researchers also point out how AI can deliver timely, relevant, and individualized feedback. Unlike traditional feedback's delay or bulkiness, AI can process feedback simultaneously as it delivers it. It is possible to cite several authors (Luan et al., 2023; Naseer et al., 2024; Vashishth et al., 2024; Yesilyurt, 2023).

There has been consensus that AI feedback generated by ChatGPT has the potential to enhance learning with its instant response to queries and submissions, though there has been a paucity of empirical evidence as to the quality improvements (Grassini, 2023; Luan et al., 2023; Michel-Villarreal, 2023). Additionally, the natural language processing (NLP) talents of ChatGPT enable it to produce context-specific responses, furnishing feedback in ways that go far beyond what "push-button" rules permit (Nepal, 2024; Rane, 2023; Wu et al., 2023). These are critical for STEM education, where students' advanced critical thinking and problem-solving should be prioritized but need ongoing and prompt feedback and assistance. Thus, AI-generated feedback in these domains can significantly benefit students by offering highly customized assistance, which human instructors often struggle to accomplish proficiently (Mollick & Mollick, 2024). Therefore, instead of replacing human feedback systems, integrating AI into educational systems becomes an extra help for the student's learning process and a source of personalized and instantaneous reactions. Moreover, it is fundamental for educational institutions to reconcile and adjust human and AI feedback to achieve an optimal equilibrium between the volume of knowledge imparted and the total cumulative learning impact (Luan et al., 2024; Nikolopoulou, 2024; Sevnarayan & Potter, 2024).

## AI AND THE ENHANCEMENT CRITICAL THINKING AND PROBLEM-SOLVING SKILLS

Acquiring analytical and problem-solving skills is an essential requirement for achieving proficiency in academic subjects. Hence, AI applications in schooling bring a lot of new prospects and at the same time challenges in fostering such skills. Engagement in critical thinking and problem-solving is a hallmark of learning, and that is central in the STEM science, technology, engineering, and mathematics (STEM) curricula, where students get multifaceted problems that demand novel answers (Ardianti et al., 2020; Baharin et al., 2018; Madden et al., 2013; Stein et al., 2007). The relationship and significance of these talents to the utilization of technical tools and AI remain unclear. The majority of the current study presents comparison data that exhibits significant variation. Reports on how digital tools and AI may affect the cultivation of these skills bring out disparate findings. For instance, many researchers have concluded that AI can contribute to developing critical reasoning and problem-solving, while this can be achieved through a personalized learning experience and learners' engagement in complex problem-solving situations (Chi & Wylie, 2014; VanLehn, 2011).

Nevertheless, there is an abundance of research findings that are incongruous and present confounding conclusions regarding the influence of AI in education. Therefore, the remaining research interest is dedicated to examining the case of AI – ChatGPT – and its connection to the specified elements. This pertains to two aspects: the ability of ChatGPT to facilitate the expression of supplementary images and concepts alongside a person's statements, and the techniques involved in the planning process, emotional intelligence, social skills, and goal setting related to the specific case mentioned (Bozkurt et al., 2023; D'Mello et al., 2017; Dwivedi et al., 2023; Nazir & Wang, 2023). Theoretical perspectives on heuristic learning and

cognitive engagement suggest that ChatGPT can support the development of critical thinking and problem-solving skills by simulating real-world problem-solving dialogues and providing immediate, context-specific feedback (Baradel, 2023; Graesser et al., 2014; Liu et al., 2023). This interaction engages students in deep cognitive processes and encourages them to apply heuristic strategies to navigate complex problems, enhancing their problem-solving abilities (Lin et al., 2024).

Incorporating artificial intelligence (AI) in educational institutions offers remarkable opportunities for striking a chord with students' cognitive processes and fostering their analytical and problem-solving skills. For instance, with the advent of AI technologies, particularly conversational AI like ChatGPT, there is now a high likelihood that those tools will enhance heuristic learning and cognitive engagement in STEM education. It is also worth noting that these technologies still need in-depth scientific investigations to facilitate a full evaluation of their impact on learner outcomes and best practices for their utilization in educational setups.

## THE ROLE OF AI SUPPORTING KNOWLEDGE RETENTION

A significant issue of conventional teaching approaches is the way in which students forget knowledge, consequently failing to transfer it. Nevertheless, AI-enhanced methods can increase the power of influence (Abdellatif et al., 2022; Kamalov et al., 2023; Martin, 2019; Vashishth et al., 2024). The difficulties of learning new information and keeping it for the long term are well-known; studies show that students often forget a lot of what they've learned right after class (Roediger & Karpicke, 2006). The cognitive factors like encoding, storage, and retrieval processes, which are responsible for the retention of new knowledge, can be influenced by multiple factors, including the engagement of audiovisual material and usage of reinforcement methods during the instruction (Cepeda et al., 2006, Dai et al., 2021, Dai & Arnberg, 2022, Dai et al., 2023; Luan, 2021; Terry, 2017). Through spaced repetition, research in AI technologies of personalized learning, learning efficiency, and achieving high levels of information retention has been observed. Thus, hand in hand with the custom intervals, spaced repetition is geared towards sharpened memory retention (Karpicke & Roediger, 2007).

AI-based personalized learning systems include adaptive learning, which allows for tailored practice to ensure better recall (Xie et al., 2020). Through ChatGPT and its fantastic text analysis and reasoning capabilities, student retention can be further improved through personalized content and interactive activities. ChatGPT, on the other hand, speeds up the process and eliminates the uncertainty as students get immediate answers to their questions as soon as they confront them. They also do the learning through dialogue with the possibility of execution so that, in the end, the learning experience becomes very tangible (Babu & Akshara, 2024; Dinçer et al., 2023). Moreover, this approach proposes that successful feedback should provide students with information regarding their present performance and learning goals and offer guidance on bridging the gap between the two (Kerfoot et al., 2010). In conclusion, AI-powered dialog systems are among the promising methods that can be used to retain knowledge in the educational process. These personalized, interactive learning practices, which are based on cognitive principles and which students utilize, would likely increase retention rates significantly, thus swelling the entire education outcome.

## AI'S IMPACT ON COLLABORATIVE LEARNING AND GROUP DYNAMICS

Team-based learning and collaborative problem-solving are essential features of modern education. They benefit from technologies such as AI, enabling new forms of student collaboration. The field of educational psychology has recognized the importance of collaborative learning, mainly through social interdependence theory and cooperative learning. These approaches have positively impacted social interaction, knowledge construction, and various aspects of student development, including cognitive, social, and psychological domains (Johnson & Johnson, 2009). Collaborative learning allows students to interact socially, to negotiate, and to share their various and individual perspectives. In his work, Slavin (1996) has shown that all student interactions resulted in a much broader grasp and recollection of the important knowledge content. Research on AI in education reveals the advantages and disadvantages of this innovative change. AI can play this role through the precision of decision-making and shortening the steps of many different processes (Javaid et al., 2022; Rane,

2023). Nevertheless, the use of AI technology in educational settings is not without concerns like equity among students for participation and the tendency of students to rely overly on technology (Pedro et al., 2019).

The role of ChatGPT must be emphasized in a group assignment, and this fact should not be underestimated. It has the possibility to carry conversations that can illustrate how such conversations affect student-student relationships, the assignment of tasks, and the learning assistance between peers when in a group project. Another example could be ChatGPT's method of placing responsibilities according to certain fields and employing a simulated dialogue to solve disagreements (Ray, 2023). Group dynamics and collaborative learning can be improved with the use of ChatGPT's communication and feedback features (Abujaber et al., 2023; Amin, 2023; Baskara, 2023; Syahputra, 2023). In addition, by facilitating discussion and directing students to relevant resources, ChatGPT can enhance interaction and collaborative learning. ChatGPT, possesses both capabilities and difficulties when utilized in conjunction with collaborative learning. While ChatGPT can enhance collaborative work, it is important to remain cautious of these concerns to maximize the benefits of working in a group. Future research should examine the impact of AI on crowd behavior and develop strategies to enhance students' learning in a scientific setting.

### BRIDGE THE GAP: THE UNIQUE POSITION OF CHATGPT IN EDUCATIONAL FEEDBACK

Undoubtedly, AI has become part of almost every industry, including education. However, ChatGPT is peculiar in that it provides comprehensive analysis and profound responses to students. The AI and its applications are gaining popularity due to its capability to adjust with the context and provide the required content regardless of the subject area (Kalla & Smith, 2023; Kasneci et al., 2023; Tlili et al., 2023). ChatGPT's ability to analyze and produce natural language enables it to provide prompt and deeply congruent feedback with the learner's comprehension and requirements (Javaid et al., 2023; Koraishi, 2023). The flexibility of this approach guarantees individualized input, which is crucial for successful learning and self-motivation (Dai, et al., 2022; Lin, 2023; Lin & Dai, 2022; Lin et al., 2022; Ray, 2023). However, the level of feedback that human feedback can provide is limited as it is often personal, which contrasts sharply with the scalability of ChatGPT.

Even though Human feedback is unparalleled in its ability to discern variations in context, this can be a minimal method for scaling solutions. However, other AI tools may provide expansion on such surfaces; ChatGPT's multilayered and deep language model appears to take the lead in this area (Brown et al., 2020; Ray, 2023). This dichotomy of scale and ability to form a varied relationship make ChatGPT a potentially game-changing tool for schools. However, we are just at the beginning of the process of using ChatGPT in education. Also, ongoing experimentation should optimize the blending of ChatGPT in learning environments, emphasizing the bettering and enriching of the human-instruction experience. Newer improvements may involve more advanced robotics, automatic personalization, and the ability to comprehend and adjust to the learner's pattern (Hirschberg & Manning, 2015). In summary, ChatGPT's distinctive role in educational feedback underscores AI's transformational capacity in education. As AI advances, it is crucial to explore its applications to guarantee it remains a formidable ally in achieving educational excellence.

## METHODOLOGY

This research empploys the convergent parallel mixed methods design (Creswell & Clark, 2023). Thus, the amalgamated quantitative and qualitative data may be assessed, and a correlation is anticipated during the interpretation phase. Using this method, we aim to provide a comprehensive understanding of ChatGPT's effects on STEM education by considering both statistics prevalence and the personal comments of program participants. The quantitative research involves the utilization of variances in grades and the retention proxy as part of the qualitative study, and this includes the content and sentiment in the comments from the two categories: ChatGPT versus humans. Therefore, the research topic being viewed from various angles leads to an increase in validity in terms of the depth of the topic (Johnson et al., 2007).

## PARTICIPANTS

The participants' pool consisted of undergraduate students enrolled in the "Introduction to Cognitive Neuroscience" course at a research institute in eastern China. This class was delivered in English, and students completed the assignments and interaction in English as well. Out of the 120 students who were registered for the course, 106 actively participated in the study, resulting in a participation percentage of 88.3%. Participants were identified through classroom announcements and email, a process that took place during the fourth week of the Spring 2023 semester. The Institutional Review Board approved the study. All possible subjects signed an informed consent form that described the study's objective, methods, signs, risks associated with research, and their rights as research participants.

Students were encouraged to participate in the study and were informed that they would be offered course credits for their involvement. However, participation remained entirely voluntary, and subjects were assured that their participation would not be detrimental to their course success. The participants agreed to join the study, signed consent forms, and returned them to the researchers. Participants ranged in age from 18 to 25 years ($M$ = 20.7, $SD$ = 1.8). The gender distribution was 52% female ($n$ = 55) and 48% male ($n$ = 51). Academic backgrounds were diverse: 65% ($n$ = 69) from social sciences, 25% ($n$ = 26) majoring in psychology, and 15% ($n$ = 16) from computer science and other disciplines. A significant fraction of the class was engaged in social sciences, of which the psychology core curriculum comprised 25%. Other disciplines, including Computer Science (15%), comprised the remaining 35%. Participants were assigned to groups – defendants and plaintiffs – per their preferences and academic interests.

The data and information were anonymized to safeguard the privacy of the participants. Furthermore, all digital data was saved on encrypted, password-protected servers. This process of selecting participants and ethical concerns is carried out appropriately, as it is in good research in education; this ensures a varied sample that represents the group while maintaining ethical standards.

## RESEARCH TOPICS AND GROUP ROLES

In order to reduce the potential introduction of biases, the study presented a selection of 20 research subjects about cognitive neuroscience for the students to select from. These topics included the influence of music on the brain, cognitive function, neural mechanisms of memory, and the impact of sleep on cognition. Students selected their preferred topics based on their interests and academic backgrounds, ensuring that each topic had a diverse representation of student preferences.

We allowed students to self-select their group members and choose their roles within the groups (Defendant or Plaintiff). The Defendant groups were responsible for creating presentation slides based on their chosen topics, while the Plaintiff groups evaluated these presentations and provided feedback. This process promoted organic group formation, and the task design was centered on students' interests.

## PROCEDURE

We introduced the study objectives and methodology to students during the fourth week of the Spring 2023 semester on the study's objectives and methodology. They selected their role as Defendant or Plaintiff groups based on their preferences and academic interests. The lecturer presented 20 study topics related to cognitive neuroscience for the groups. The students were given relevant material for each research topic, including references and critical arguments, as well as the tasks that they were anticipated to accomplish. Defendant groups were required to create the presentation slides based on their selected topics and strategies, while Plaintiff groups evaluated these presentations and provided suggestions for improvement. Evaluation criteria were standardized and disseminated to all groups beforehand to ensure fairness and objectivity. Group members were self-selected, promoting organic group formation, with the lecturer approving the final team compositions.

Defendant and Plaintiff groups were required to submit their work to the lecturer via email in a specific order. First, Defendant groups were required to submit their presentation slides, which the lecturer then graded (this constitutes their pre-feedback score). The lecturer sent human or

AI feedback to the Defendant group according to the random assignment results. Subsequently, the lecturer forwarded the Defendant group's slides to the corresponding Plaintiff groups. The Plaintiff groups were given one week to review the slides, identify issues in the arguments, and provide their feedback. After one week, the Plaintiff groups were required to submit their feedback to the lecturer, who then graded it (this constitutes the Plaintiff group's pre-feedback score). The lecturer sent human or AI feedback to the Plaintiff groups based on the random assignment results.

Both groups were given one week to revise their work based on the feedback they received. After this period, the Defendant groups needed to upload their revised presentation slides to Canvas. The corresponding Plaintiff groups were required to upload their final work, which incorporated the feedback (human or AI), as a post under the corresponding Defendant group's submission. The lecturer then re-evaluated the updated work of both groups using the same evaluation criteria, assigning post-feedback scores.

## FEEDBACK ASSIGNMENT AND PRESENTATION PREPARATION

Participants in the Defendant and Plaintiff groups were randomly assigned to receive either human feedback from the lecturer or AI-generated feedback from ChatGPT. This randomization aimed to minimize selection bias and bolster the study's internal validity.

## RESEARCH QUESTIONS

The study's analysis was structured to examine three research questions:

- RQ1: How does ChatGPT feedback influence the development of critical thinking and problem-solving skills in students compared to traditional human feedback? This question addresses the potential of ChatGPT to foster analytical thinking and the application of knowledge in problem-solving scenarios, which are crucial skills in STEM education.
- RQ2: What is the effect of ChatGPT feedback on the long-term retention of course material? Knowledge retention is a significant challenge in education, and this question explores whether the immediacy and personalization of ChatGPT feedback can lead to more durable learning outcomes.
- RQ3: How does the integration of ChatGPT into group projects affect collaborative learning dynamics and outcomes? Given the importance of teamwork and communication in higher education, this question examines the role of ChatGPT in facilitating and enhancing group interactions and learning processes.

## ANALYSIS PLAN

We structured our analysis into three main components: grade change analysis (grade change = post-feedback score – pre-feedback score), retention proxy analysis, and collaborative learning dynamics analysis.

## MEASURES

### 1. GRADE CHANGE ANALYSIS

- *Variables*: Pre- and post-feedback grades for each student group were used to calculate grade changes (grade change = post-feedback score – pre-feedback score). Feedback type (ChatGPT vs. Human) was categorized.
- *Measure*: Grade change is calculated as the difference between pre-and post-feedback.
- *Data Collection*: pre-feedback and post-feedback scores provided by the lecturer were collected. The score in the pre-feedback assessment was considered as the baseline.
- *Statistical Analysis*: Our method used the Wilcoxon rank-sum test to compare the differences in median grades between ChatGPT and the human groups). This non-parametric test was selected based on its capacity to compare two independent groups without assuming a normal distribution (Fay & Proschan, 2010) recommended. The Wilcoxon rank-sum test ranks all values that reflect grade changes in both groups, thus comparing the sum of ranks for each group to get the results. The Mann-Whitney U tests are highly sensitive statistical tests for evaluating the effectiveness of this strategy. If the

ranks differ, it would be considered an unexpected discrepancy, which might serve as quantitative evidence (Mann & Whitney, 1947).

### 2. RETENTION PROXY ANALYSIS

The Retention Proxy is a composite score derived from a survey that evaluates students' perceptions of the feedback they received. It is designed to measure factors related to long-term retention of course material. Here is a breakdown of how it was calculated:

- *Variables*: The Retention Proxy score was composed of five variables:
  1. General Experience
  2. Feedback Usefulness
  3. Relevance
  4. Engagement
  5. Motivation
- *Measure*: The Retention Proxy score was derived from a 7-point Likert scale survey (1 = Very Poor, 7 = Excellent) for each variable.
- *Data Collection*: Each group completed the survey assessing these variables immediately after receiving feedback at the end of the semester.
- *Statistical Analysis*: Descriptive statistics (mean, median, standard deviation) were calculated for Retention Proxy scores. The Wilcoxon rank-sum test was used to compare these scores between feedback types. Cohen's d was computed to measure the effect size of differences (Cohen, 2013). All variables in the Retention Proxy score were equally weighted in the composite score calculation. This approach was based on previous literature, which deemed each aspect equally important to the retention construct (Karpicke & Roediger, 2007). The Retention Proxy score was calculated as the average of the five variables:

$$\text{Retention Proxy} = \frac{\text{General Experience} + \text{Feedback Usefulness} + \text{Relevance} + \text{Engagement} + \text{Motivation}}{5}$$

### 3. COLLABORATIVE LEARNING DYNAMICS ANALYSIS

- *Variables*: To understand the content and tone of feedback, feedback texts were subjected to word count, top verb analysis, sentiment analysis (using AFINN), and top bigram (N-gram) analysis. Additional feedback was the primary text variable analyzed.
- *Measure*: AFINN sentiment analysis (Nielsen, 2011), word frequency counts, and bigram analysis.
- *Data Collection*: Each group completed the survey and provided their feedback about the feedback they received at the end of the semester
- *Statistical Analysis*:
  - Word Count & Top Verbs: Descriptive analysis quantified the frequency of words and identified the top 10 verbs used in each feedback type.
  - Sentiment Analysis: AFINN sentiment scores were calculated for each feedback type, and mean sentiment scores were compared.
  - Top Bigrams (N-gram) Analysis: Identified and counted the frequency of bigrams in the feedback texts, comparing the top 10 for each feedback type.

### DATA ANALYSIS SOFTWARE AND TOOLS

- The analysis was conducted using R, employing packages such as dplyr for data manipulation, ggplot2 for visualization, rstatix and report for statistical testing, and APA-style reporting.

## RESULTS

### RESEARCH QUESTION 1: IMPACT OF FEEDBACK TYPE ON GRADE CHANGES

A Wilcoxon rank-sum test was conducted to compare the change in grades between students who received feedback from ChatGPT and those who received human feedback. The results indicated a statistically significant difference in the change in grades between the ChatGPT

feedback group (median = −12) and the human feedback group (median = −5), $W = 5.5$, $p < .001$, with both groups having 10 participants each. This suggests that the type of feedback received significantly impacts students' grade improvement, with students who received ChatGPT feedback experiencing a more significant median decline in grades than those who received human feedback. The negative median grade changes for both groups suggest that, on average, grades declined from the initial to the final assessment within both the ChatGPT and Human feedback groups, as shown in Figure 1. However, the decline was more pronounced in the ChatGPT group. The statistically significant result indicates that the difference in grade changes between the two feedback types is not likely due to random chance. The larger decrease in grades among students who received ChatGPT feedback may suggest issues related to the effectiveness of the feedback, its relevance, or how it was understood and acted upon by the students. It can reflect differences in the quality or applicability of feedback or how personalized and actionable the feedback was compared to human-provided feedback.

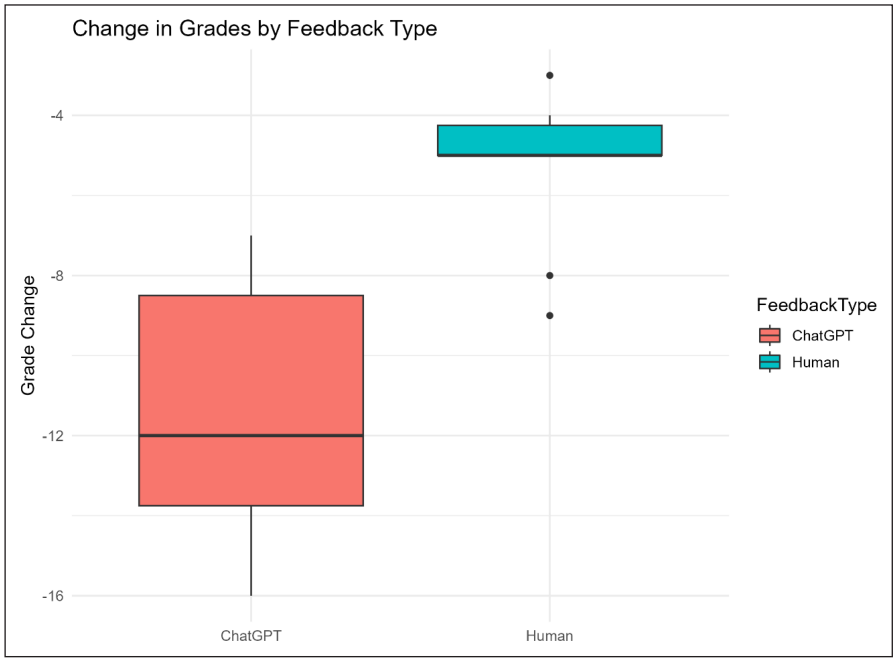


**Figure 1** Grades Change by Feedback Type.

## RESEARCH QUESTION 2: INFLUENCE OF FEEDBACK TYPE ON LONG-TERM RETENTION

The Retention Proxy measure, composed of equally weighted scores for General Experience, Feedback Usefulness, Relevance, Engagement, and Motivation, provides insight into factors contributing to long-term course material retention. Our findings show that human feedback ($M = 5.78$, $SD = 0.405$) was associated with significantly higher Retention Proxy scores than ChatGPT feedback ($M = 4.06$, $SD = 0.411$). A Wilcoxon rank-sum test indicated a statistically significant difference between the Retention Proxy scores for the two feedback types, $W = 0.5$, $p < .0001$. The effect size, measured by Cohen's d, was substantial ($d = -4.22$, 95% CI [−5.34, −3.08]), indicating that human feedback was associated with a significantly higher Retention Proxy score than ChatGPT feedback. This suggests that human feedback may be more effective in promoting long-term retention, possibly due to its ability to provide nuanced, contextually relevant guidance that resonates more deeply with students, as shown in Figure 2.

These findings support the hypothesis that the feedback received can significantly influence factors related to the long-term retention of course material. Human feedback shows a more pronounced positive impact.

## RESEARCH QUESTION 3: INTEGRATION OF CHATGPT IN COLLABORATIVE LEARNING

The sentiment analysis results indicate a higher net sentiment score for feedback categorized under "Human" (sentiment score = 63) compared to "ChatGPT" (sentiment score = 40). This

suggests that human-generated feedback is more positive or less negative than feedback associated with ChatGPT, which can foster a more positive collaborative environment, as shown in Figures 3 and 4.

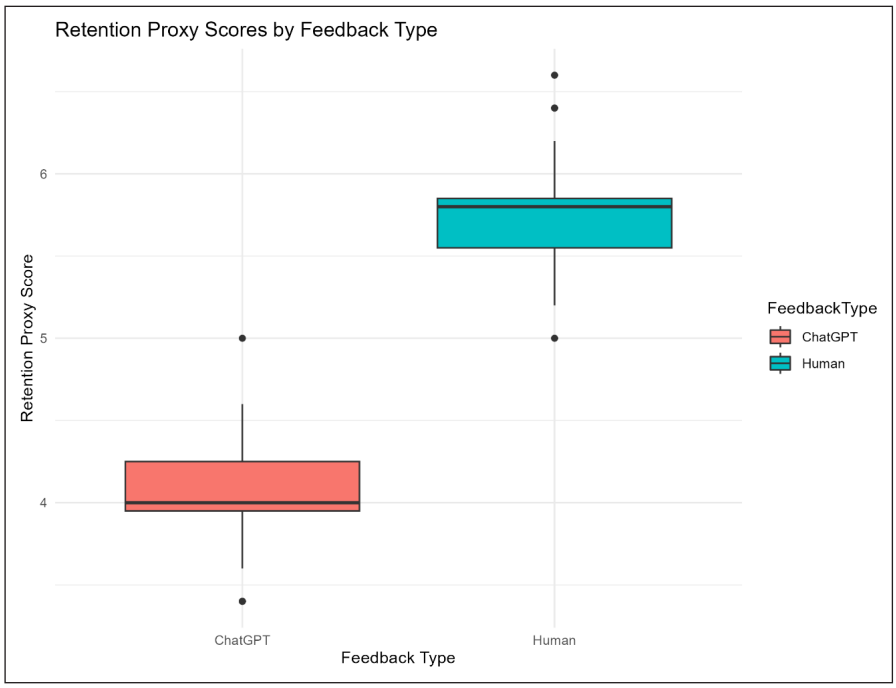


**Figure 2** Retention Proxy Scores by Feedback Type.

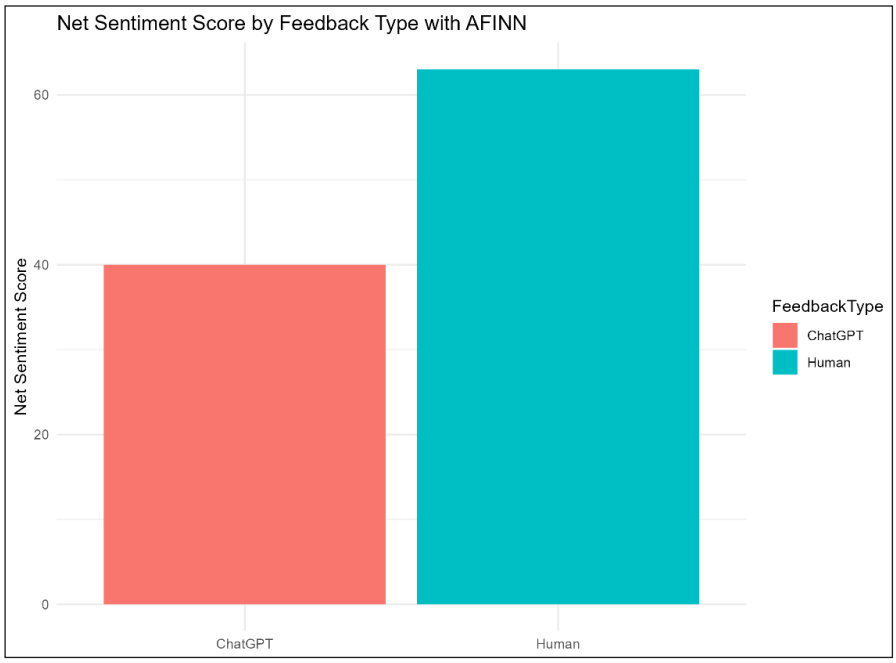


**Figure 3** Sentiment Score by Feedback Type with AFINN.

The word count results show that "Human" feedback contains more words (939) than "ChatGPT" feedback (610). This discrepancy might suggest that human participants provide more detailed feedback than ChatGPT, which can be crucial for learning as it may offer more specific guidance and insights, potentially leading to better understanding and improvement in group projects. Specifically, the human feedback group students perceive feedback as being more valuable as it offers comprehensive details and precise assistance for their work, and more details were recalled and provided in the open-ended question in Figure 5.

In Table 1, we compare the main ten verbs for every feedback type received from people and ChatGPT; Table 1 highlights the distinctions in the actions emphasized by the two. Examples of these are "add," "make," and "understand," which are the most frequently used in human feedback, while "be," "explore," and "lead" are the most frequently used in ChatGPT feedback. The dissimilarity in the verbs used for remarks can indicate a focus on slightly different things: human feedback can be more directed at improvements and even understanding ideas, while

ChatGPT can be more centered on exploration and leadership. Such differences may lead to a more apparent distinction in the dynamics of collaborative learning, whereby the responses of individual group members vary depending on the status of the feedback.

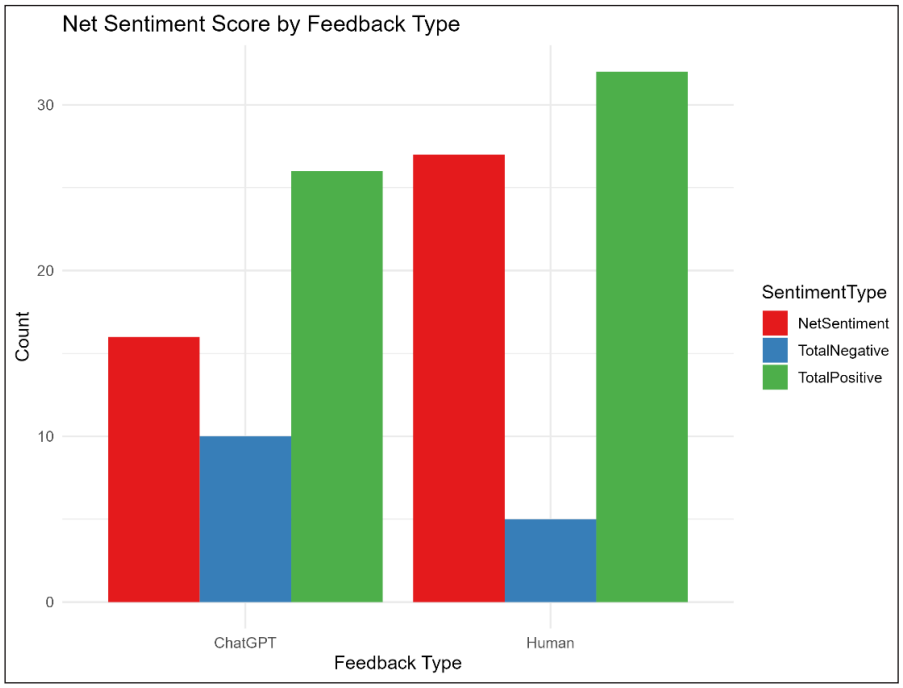


**Figure 4** Sentiment Score by Feedback Type.

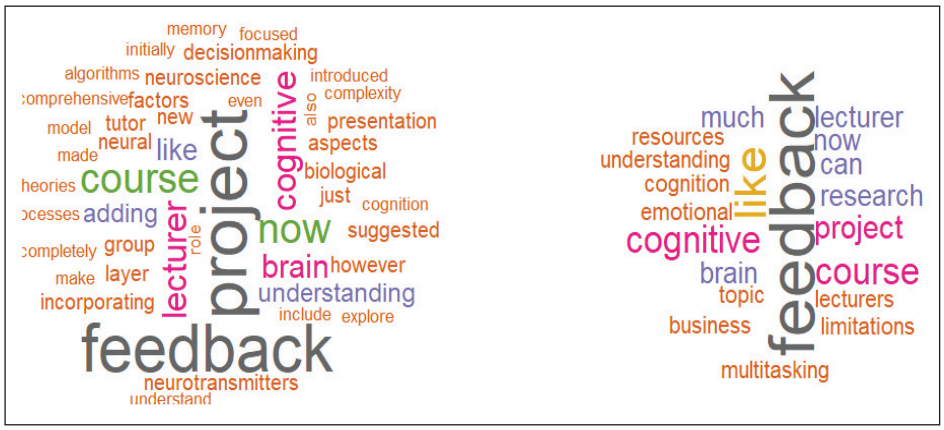


**Figure 5** Wordcloud for Human Feedback Group and ChatGPT Feedback Group.

*Note:* The figure on the left represents the keywords in open-ended questions for the Human feedback group, and the figure on the right represents the keywords from the ChatGPT group.

| | ChatGPT | Human |
|---|---|---|
| 1 | be | add |
| 2 | explore | make |
| 3 | lead | understand |
| 4 | add | be |
| 5 | apply | focus |
| 6 | consider | include |
| 7 | enrich | explore |
| 8 | feel | incorporate |
| 9 | focus | suggest |
| 10 | have | think |

**Table 1** Top 10 Verbs for ChatGPT and Human Feedback Group.

The analysis of the top 10 N-grams (bigrams) within each group provides detailed insights into the themes and topics emphasized in the feedback, as shown in Figure 6. This linguistic analysis, particularly focusing on verbs, is crucial as verbs often indicate the type of cognitive processes and actions encouraged in educational feedback. Verbs are crucial in educational feedback,

and the lack of English language proficiency will limit the understanding of these cognitive processes and actions in generating feedback (Biber & Conrad, 2019; Hyland & Tse, 2004). In the case of feedback generated by ChatGPT for groups labeled "lecturers feedback" or "cognitive functions," it is likely of educational type feedback and cognitive processes aspect of the learning. The use of action verbs like "explore" and "lead" in ChatGPT feedback (Table 1) shows a preference for student-centered learning by incorporating the principles of active exploration and inquiry. This aligns with contemporary educational theories that put students in the center of the learning process (Weimer, 2013). Differently, human-generated feedback classes having bigrams, such as, "cognitive neuroscience" and "biological aspects," provide evidence of the more scientific or detailed questioning. The overuse of verbs such as "add," "make," or/and "understand" in human feedback, as just example, can mean that the focus is on empirical upheaval or bereft understanding. This is quite the opposite of the student-centered learning technique, which majors on knowledge building and use (Anderson & Krathwohl, 2001). The variance in theme and structure and grammar of the containing contained in feedback would importantly, among other things, divert attention from what students should concentrate on in their projects. Employing such verbs in feedback can affect the way students engage in any cognitive activities and approach learning tasks.

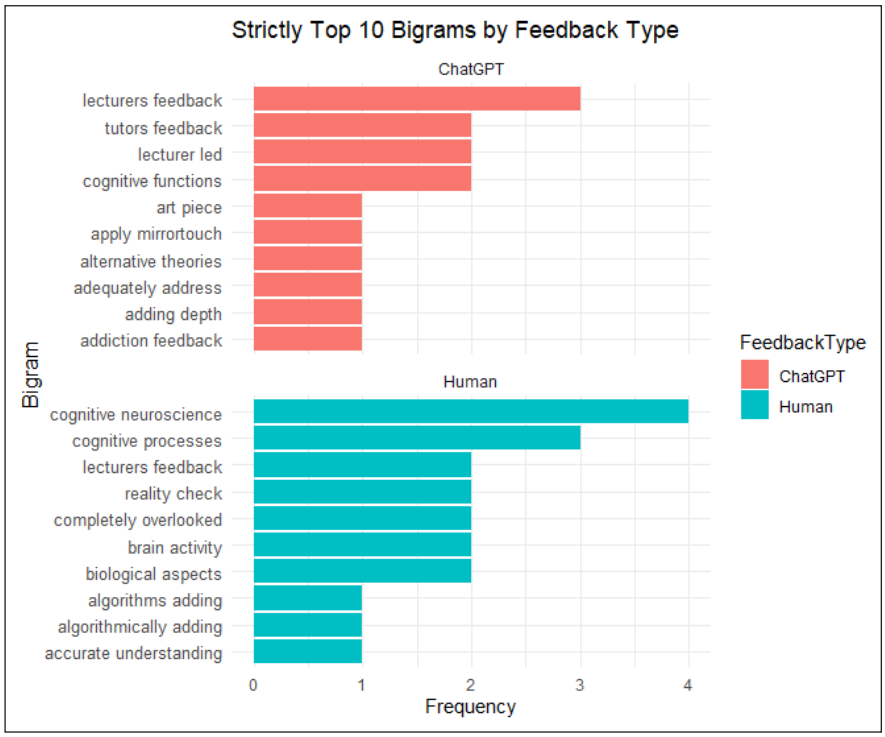


**Figure 6** Strictly Top 10 Bigrams by Feedback Type.

Consequently, the incorporation of ChatGPT into group projects modifies the way the collaborative learning interactions and outcomes are formed in several dimensions. When compared with ChatGPT's data, human feedback is generally characterized by a greater degree of positivity, detail, and focus on enhancing understanding and students' performance. The thematic emphasis of feedback also diverges between humans and ChatGPT, which might draw the attention and focus of team members in different directions. These results reveal that although human feedback is still unreplaceable for creating an accessible, comprehensive, and progress-oriented learning environment, ChatGPT is a powerful teaching-and-learning assistant tool for group feedback delivery. We suggest future research delve deeper into the approaches in which these variations in feedback affect particular learning outcomes and group interaction processes.

## DISCUSSION AND CONCLUSION

The study analyzes and compares AI-generated and lecturer-led feedback on students' learning goals in higher education. We focus on grade transformations, long-term retention of course materials, and collaboration. The results reveal that AI-generated feedback is effective in educational contexts. The findings contribute to the continuing discourse on AI as an effective educational tool (Kasneci et al., 2023; Tlili et al., 2023).

### GRADE CHANGES AND CRITICAL THINKING DEVELOPMENT

The study found that the ChatGPT feedback group's grades statistically significantly improved compared with the human feedback group. The students in the ChatGPT feedback group exhibited a much more significant decline in their grades (median = −12) than those receiving human feedback (median = −5). These research results contradict the theory that any personalized feedback can be duplicated by an AI program in classroom (Luan et al., 2023). However, the high reduction in grades of the AI ChatGPT group indicates that the AI-generated feedback might not be effective when it comes to the tasks that require critical thinking and problem-solving skills, which are essential in STEM education (Ardianti et al., 2020; Baharin et al., 2018). Although the study suggests that AI improves learning outcomes in many ways, human instructors are still unreplaceable in individualized guidance (Grassini, 2023; Michel-Villarreal, 2023).

### LONG-TERM RETENTION AND FEEDBACK QUALITY

Human feedback produced superior retention scores compared to ChatGPT feedback by a substantial margin. Meanwhile, a considerable Cohen's effect size (d = −4.22), supporting the view that human feedback may be more effective in fostering the long-term retention of learning materials, was detected. This corroborates earlier studies with simple or direct evidence on learning retention and consolidation of memory due to affective feedback (Karpicke & Roediger, 2007; Xie et al., 2020). The human feedback conveyed the information in the most engaging way, and the learning remained deep and memorable. This is because it is more adaptive, specific, and relatable advice that connects better with the learners than generic automated feedback. This concept aligns with the heuristic learning and cognitive engagement learning processes, which posit that individualized and interactive educational experiences can substantially enhance recall rates (Graesser et al., 2014; Liu et al., 2023).

### COLLABORATIVE LEARNING DYNAMICS

The experiments that were conducted on the applications of ChatGPT in group projects showed a variety of collaborative learning dynamics. On the aforementioned score, our analysis exhibited a better sentiment score sent by human-generated feedback (63), which was way higher than that one sent by ChatGPT feedback (40), signaling a more affirmative sentiment. Furthermore, human feedback encompassed a more significant number of words (939) than ChatGPT feedback (610), which is a norm in human-to-human communication. The examination of verb and bigram usage proved the source of knowledge in the two systems, human and artificial intelligence. The human approach incorporates more verbs, such as "add," "make," and "understand," whilst ChatGPT only uses "explore" and "lead." Given the emerging evidence that students may notice these pattern differences as differences between AI writing and human writing, and they may stake their response mostly to this pattern, the implications of these findings are that while ChatGPT may serve as the primary means for providing feedback in collaborative learning settings, it lacks the capacity to foster a positive or optimistic environment and be detailed or problem-centric as human feedback. The outcome has affirmed the proposition that it is supposed to be a combinational approach for tasking purposes where both AI and the human mind are much needed to optimize learning dynamics and the set Sharif outcomes of light on (Adıgüzel et al., 2023; Rasul et al., 2023).

## LIMITATION AND FUTURE RESEARCH

The findings of this study illustrated that such AI interventions are indeed not easy. Concerning ChatGPT, the new model of delivering consistent and scalable feedback is promising; however, the results give us insight into the fact that the model still trails behind human feedback in encouraging fast grade improvement, long-term retention, and a conducive collaborative learning environment. Besides this study's limitations, it also includes the small sample size and the investigation's context, which might impact the generalizability of findings. COVID-19 has been an "icebreaker" in this fast transition of the education hypotheses, but it also poses some difficulties for teachers (Luan et al., 2023). Potential studies postulate the following topics of future studies: the number of words in the feedback, interactions, and how feedback contribution influences retention and learning

outcomes. Such research should further introduce students' "English proficiency" as one of the confounding factors that may have a negative effect on the study outcome. On top of that, research spanning remote populations and different instructional forums will give even better answers to the effectiveness of ChatGPT-generated academic feedback. Personalized feedback improves student learning outcomes; thus, it should not only focus on written feedback but also on students receiving oral feedback on presentation skills and complex math problems (one of the methods through which this was done was by sending them to various talks that discussed topics they studied). The second study on the above AI technologies also proved that they can increase the efficiency of feedback or generate subject-based feedback, such as mathematics, for the students' variations (Lei & Xin, 2023; Lei et al., 2020a; Lei et al., 2020b) and conduct convergence studies to empower education with AI (Guo & Wang, 2024; Wang et al., 2023). We recognize that the skills or abilities of the prompt itself used in generating ChatGPT feedback can impact the IMI results as well. Although the prompts we chose represent a deliberate process, it is fascinating to check how further prompt development can improve its performance.

In conclusion, this study highlights the way that different types of feedback—ChatGPT versus human—affect students' learning outcomes. While ChatGPT-generated feedback presents a promising tool for educational feedback, human feedback's unique contributions to fostering a conducive learning environment cannot be overlooked. This study would provide a baseline for comparing AI-generated feedback and human feedback. Further investigations into the optimal prompt design can potentially narrow the gap in effectiveness between AI-generated and human feedback. Moreover, the findings would contribute to the ongoing discourse on integrating AI tools in education and underscore the importance of blending human and AI-generated feedback to optimize learning dynamics and outcomes.

## APPENDIX

Prompt for Defendant Groups:" You act as a lecturer in cognitive neuroscience, particularly knowledgeable about [Topic #]. You have been asked to review a presentation slide on this topic created by undergraduate students, who are the Defendant (the Defendant group needs to select one point of view and build up the argument to support their argument based on the previous studies). Please provide detailed, constructive feedback on the following aspects:

1. Content accuracy: Evaluate the scientific accuracy of the information presented. Identify any misconceptions or areas where the latest research findings are not reflected.
2. Depth of analysis: Assess whether the presentation covers the topic in sufficient depth for an introductory cognitive neuroscience course. Suggest areas where more detail or explanation might be beneficial.
3. Clarity of explanations: Comment on how well complex concepts are explained. Offer suggestions for improving clarity where needed.
4. Use of evidence: Evaluate how well the presentation incorporates and cites relevant research. Suggest additional key studies or data that can strengthen the argument.
5. Structure and flow: Assess the logical flow of ideas in the presentation. Suggest improvements to the structure if necessary.
6. Visual aids: Comment on the effectiveness of any diagrams, charts, or other visual elements used. Suggest improvements or additional visuals that can enhance understanding.
7. Strengths: Highlight the strongest aspects of the presentation.
8. Areas for improvement: Identify the most critical areas where the presentation can be enhanced, providing specific suggestions for each.

Please provide your feedback in a constructive manner, balancing critique with positive reinforcement to encourage the students' learning and improvement."

Prompt for Plaintiff Groups: "You act as a lecturer in cognitive neuroscience, particularly knowledgeable about [Topic #]. You have been asked to review feedback provided on the student presentation about this topic (the feedback provided by the Plaintiff group, whose task is to evaluate the presentation slide from content and structure as well as the visual presentation approach. Your task is to evaluate the quality and comprehensiveness of this feedback. Please address the following points:

1. Accuracy of critique: Assess whether the feedback accurately identifies strengths and weaknesses in the presentation's content. Note any misunderstandings or overlooked issues in the feedback.
2. Constructiveness: Evaluate how well the feedback balances critique with encouragement. Suggest ways to make the feedback more constructive if necessary.
3. Specificity: Determine if the feedback provides specific, actionable suggestions for improvement. Identify areas where more detailed guidance can be beneficial.
4. Comprehensiveness: Assess whether the feedback addresses all key aspects of the presentation (content, structure, visuals, etc.). Suggest additional areas that should be covered in the feedback.
5. Scientific rigor: Evaluate how well the feedback reflects current scientific understanding of [Topic #]. Suggest any additional key points or recent research that should be mentioned.
6. Clarity: Comment on how clearly the feedback is communicated. Suggest improvements in explanation or language use if needed.
7. Potential impact: Assess how likely this feedback is to lead to significant improvements in the presentation. Suggest ways to enhance the feedback's potential impact on the students' learning.
8. Overall quality: Provide an overall assessment of the feedback's quality and usefulness for improving the presentation on [Topic #].

Please provide your evaluation in a detailed, analytical manner, considering both the strengths and potential improvements of the feedback given."

## DATA ACCESSIBILITY STATEMENT

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

## ETHICS AND CONSENT

IRB Ethical approval was obtained for the work described in this article.

## COMPETING INTERESTS

The authors have no competing interests to declare.

## AUTHOR NOTES

This paper was proofread, edited, and refined with the assistance of OpenAI's GPT-4o complementing the human editorial process. The human author critically assessed and validated the content to maintain academic rigor. The author also assessed and addressed potential biases inherent in AI-generated content. The final version of the paper is the sole responsibility of the human author (adapted from Bozkurt, 2024).

## AUTHOR CONTRIBUTIONS (CRediT)

Lingfei Luan: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing; Xi Lin: Conceptualization, Data curation, Formal Analysis, Writing – original draft, Writing – review & editing; Yan Dai: Validation, Visualization, Writing – original draft, Writing – review & editing. All authors have read and agreed to the published version of the manuscript.

## AUTHOR AFFILIATIONS

**Lingfei Luan** orcid.org/0000-0002-8574-6578
Psychological Science, Belmont University, USA

**Xi Lin** orcid.org/0000-0003-2387-4117
Department of Interdisciplinary Professions, East Carolina University, USA

**Yan Dai** orcid.org/0000-0003-1170-7776
Department of Educational Foundations, Leadership, and Technology (EFLT), Auburn University, USA

## REFERENCES

**Abdellatif, H., Al Mushaiqri, M., Albalushi, H., Al-Zaabi, A. A., Roychoudhury, S.,** & **Das, S.** (2022). Teaching, learning and assessing anatomy with artificial intelligence: The road to a better future. *International Journal of Environmental Research and Public Health, 19*(21), 14209. https://doi.org/10.3390/ijerph192114209

**Abujaber, A. A., Abd-Alrazaq, A., Al-Qudimat, A. R., Nashwan, A. J.,** & **AbuJaber, A.** (2023). A strengths, weaknesses, opportunities, and threats (SWOT) analysis of ChatGPT integration in nursing education: A narrative review. *Cureus, 15*(11). https://doi.org/10.7759/cureus.48643

**Acemoglu, D.** (2002). Technical change, inequality, and the labor market. *Journal of Economic Literature, 40*(1), 7–72. https://doi.org/10.1257/0022051026976

**Adıgüzel, T., Kaya, M. H.,** & **Cansu, F. K.** (2023). Revolutionizing education with AI: Exploring the transformative potential of ChatGPT. *Contemporary Educational Technology*. https://doi.org/10.30935/cedtech/13152

**Ahmad, K., Qadir, J., Al-Fuqaha, A., Iqbal, W., El-Hassan, A., Benhaddou, D.,** & **Ayyash, M.** (2020). Data-driven artificial intelligence in education: A comprehensive review. https://doi.org/10.35542/osf.io/zvu2n

**Amin, M. Y. M.** (2023). AI and chat GPT in language teaching: Enhancing EFL classroom support and transforming assessment techniques. *International Journal of Higher Education Pedagogies, 4*(4), 1–15. https://doi.org/10.33422/ijhep.v4i4.554

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

**Ardianti, S., Sulisworo, D., Pramudya, Y.,** & **Raharjo, W.** (2020). The impact of the use of STEM education approach on the blended learning to improve student's critical thinking skills. *Universal Journal of Educational Research, 8*(3), 24–32. https://doi.org/10.13189/ujer.2020.081503

**Autor, D. H.** (2015). Why are there still so many jobs? The history and future of workplace automation. *Journal of Economic Perspectives, 29*(3), 3–30. https://doi.org/10.1257/jep.29.3.3

**Babu, C. S.,** & **Akshara, P. M.** (2024). Revolutionizing conversational AI: unleashing the power of ChatGPT-Based applications in generative AI and natural language processing. In *Advanced Applications of Generative AI and Natural Language Processing Models* (pp. 228–248). IGI Global. https://doi.org/10.4018/979-8-3693-0502-7.ch011

**Baharin, N., Kamarudin, N.,** & **Manaf, U. K. A.** (2018). Integrating STEM education approach in enhancing higher order thinking skills. *International Journal of Academic Research in Business and Social Sciences, 8*(7), 810–821. https://doi.org/10.6007/ijarbss/v8-i7/4421

**Baradel, C.** (2023). Interaction, Design, and Assessment: An Exploratory Study on ChatGPT in Language Education. http://dspace.unive.it/handle/10579/25711

**Baskara, F. R.** (2023). AI-Driven dynamics: ChatGPT transforming ELT teacher-student interactions. *Lensa: Kajian Kebahasaan, Kesusastraan, dan Budaya, 13*(2), 261–275. https://doi.org/10.26714/lensa.13.2.2023.261-275

**Bhutoria, A.** (2022). Personalized education and artificial intelligence in the United States, China, and India: A systematic review using a human-in-the-loop model. *Computers and Education: Artificial Intelligence, 3*, 100068. https://doi.org/10.1016/j.caeai.2022.100068

**Biber, D.,** & **Conrad, S.** (2019). *Register, genre, and style*. Cambridge University Press. https://doi.org/10.1017/9781108686136

**Bozkurt, A.** (2024). GenAI et al.: Cocreation, Authorship, Ownership, Academic Ethics and Integrity in a Time of Generative AI. *Open Praxis, 16*(1), 1–10. https://doi.org/10.55982/openpraxis.16.1.654

**Bozkurt, A., Xiao, J., Lambert, S., Pazurek, A., Crompton, H., Koseoglu, S., Farrow, R., Bond, M., Nerantzi, C., Honeychurch, S., Bali, M., Dron, J., Mir, K., Stewart, B., Costello, E., Mason, J., Stracke, C. M., Romero-Hall, E., Koutropoulos, A., Toquero, C. M., Singh, L Tlili, A., Lee, K., Nichols, M., Ossiannilsson, E., Brown, M., Irvine, V., Raffaghelli, J. E., Santos-Hermosa, G Farrell, O., Adam, T., Thong, Y. L., Sani-Bozkurt, S., Sharma, R. C., Hrastinski, S.,** & **Jandrić, P.** (2023). Speculative futures on ChatGPT and generative artificial intelligence (AI): A collective reflection from the educational landscape. *Asian Journal of Distance Education, 18*(1), 53–130. https://doi.org/10.5281/zenodo.7636568

**Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ...** & **Amodei, D.** (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems, 33*, 1877–1901. https://doi.org/10.18653/v1/2022.emnlp-main.90

**Brynjolfsson, E.,** & **McAfee, A.** (2011). *Race against the machine: How the digital revolution is accelerating innovation, driving productivity, and irreversibly transforming employment and the economy*. Brynjolfsson and McAfee.

**Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T.,** & **Rohrer, D.** (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin, 132*(3), 354. https://doi.org/10.1037/0033-2909.132.3.354

**Chi, M. T.,** & **Wylie, R.** (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist, 49*(4), 219–243. https://doi.org/10.1080/00461520.2014.965823

**Cohen, J.** (2013). *Statistical power analysis for the behavioral sciences*. routledge. https://doi.org/10.4324/9780203771587

**Creswell, J. W.,** & **Plano Clark, V. L.** (2023). Revisiting mixed methods research designs twenty years later. *Handbook of mixed methods research designs*, 21–36. https://doi.org/10.4135/9781529614572.n6

**Dai, Y.,** & **Arnberg, B.** (2022). "We have to survive, first": Speculative ethnographies of Chinese student experience during COVID-19. *Cultural Studies↔ Critical Methodologies, 22*(1), 53–65. https://doi.org/10.1177/15327086211050041

**Dai, Y., Lin, X.,** & **Li, L.** (2021). Technology acceptance of LMS—do previous online learning experiences matter?. *Journal of Educational Technology Development and Exchange (JETDE), 14*(2), 75–90. https://doi.org/10.18785/jetde.1402.04

**Dai, Y. Lin, X., Su, S.,** & **Li, L.** (2022). The Online Learning Academic Achievement of Chinese Students during the COVID-19 Pandemic: The Role of Self-Regulated Learning and Academic Entitlement. *International Journal of Psychology and Educational Studies*, *8*(3), 116–127. https://doi.org/10.52380/ijpes.2021.8.3.384

**Dai, Y., Luan, L.,** & **Lin, X.** (2023). The effects of online learning readiness on self-regulated learning for the first-time online learning students. *Asian Journal of Distance Education, 18*(2), 42–62. https://doi.org/10.5281/zenodo.8194076

**Dinçer, R., Dinçer, N.,** & **Guksu, O.** (2023). An interactive conversation with a chatbot: Does ChatGPT know standard phraseology in aviation English? *The Literacy Trek, 9*(2), 24–41. https://doi.org/10.47216/literacytrek.1382543

**D'Mello, S., Dieterle, E.,** & **Duckworth, A.** (2017). Advanced, analytic, automated (AAA) measurement of engagement during learning. *Educational Psychologist, 52*(2), 104–123. https://doi.org/10.1080/00461520.2017.1281747

**Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., ...** & **Wright, R.** (2023). "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management, 71*, 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

**Fay, M. P.,** & **Proschan, M. A.** (2010). Wilcoxon-Mann-Whitney or t-test? On assumptions for hypothesis tests and multiple interpretations of decision rules. *Statistics Surveys*, *4*, 1. https://doi.org/10.1214/09-ss051

**Felder, R. M.,** & **Brent, R.** (2024). *Teaching and learning STEM: A practical guide*. John Wiley & Sons.

**Ghandour, A., Woodford, B. J.,** & **Abusaimeh, H.** (2024). Ethical considerations in the use of ChatGPT: An exploration through the lens of five moral dimensions. *IEEE Access*. https://doi.org/10.1109/access.2024.3394243

**Gill, S. S., Xu, M., Patros, P., Wu, H., Kaur, R., Kaur, K., ...** & **Buyya, R.** (2024). Transformative effects of ChatGPT on modern education: Emerging era of AI chatbots. *Internet of Things and Cyber-Physical Systems, 4*, 19–23. https://doi.org/10.1016/j.iotcps.2023.06.002

**Gligorea, I., Cioca, M., Oancea, R., Gorski, A. T., Gorski, H.,** & **Tudorache, P.** (2023). Adaptive learning using artificial intelligence in e-learning: a literature review. *Education Sciences, 13*(12), 1216. https://doi.org/10.3390/educsci13121216

**Graesser, A. C., Li, H.,** & **Forsyth, C.** (2014). Learning by communicating in natural language with conversational agents. *Current Directions in Psychological Science, 23*(5), 374–380. https://doi.org/10.1177/0963721414540680

**Grassini, S.** (2023). Shaping the future of education: exploring the potential and consequences of AI and ChatGPT in educational settings. *Education Sciences, 13*(7), 692. https://doi.org/10.3390/educsci13070692

**Guo, S.,** & **Wang, Y.** (2024). Investigating predictors of juvenile traditional and/or cyber offense using machine learning by constructing a decision support system. *Computers in Human Behavior, 152*, 108079. https://doi.org/10.1016/j.chb.2023.108079

**Hattie, J.,** & **Timperley, H.** (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112. https://doi.org/10.3102/003465430298487

**Hirschberg, J.,** & **Manning, C. D.** (2015). Advances in natural language processing. *Science, 349*(6245), 261–266. https://doi.org/10.1126/science.aaa8685

**Hyland, K.,** & **Tse, P.** (2004). Metadiscourse in academic writing: A reappraisal. *Applied Linguistics*, *25*(2), 156–177. https://doi.org/10.1093/applin/25.2.156

**Javaid, M., Haleem, A., Singh, R. P., Khan, S.,** & **Khan, I. H.** (2023). Unlocking the opportunities through ChatGPT Tool towards ameliorating the education system. *BenchCouncil Transactions on Benchmarks, Standards and Evaluations, 3*(2), 100115. https://doi.org/10.1016/j.tbench.2023.100115

**Javaid, M., Haleem, A., Singh, R. P.,** & **Suman, R.** (2022). Artificial intelligence applications for industry 4.0: A literature-based study. *Journal of Industrial Integration and Management, 7*(01), 83–111. https://doi.org/10.1142/s2424862221300040

**Johnson, D. W.,** & **Johnson, R. T.** (2009). An educational psychology success story: Social interdependence theory and cooperative learning. *Educational Researcher, 38*(5), 365–379. https://doi.org/10.3102/0013189X09339057

**Johnson, R. B., Onwuegbuzie, A. J.,** & **Turner, L. A.** (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research, 1*(2), 112–133. https://doi.org/10.1177/1558689806298224

**Kalla, D.,** & **Smith, N.** (2023). Study and analysis of chat GPT and its impact on different fields of study. *International Journal of Innovative Science and Research Technology, 8*(3). https://ssrn.com/abstract=4402499

**Kamalov, F., Santandreu Calonge, D.,** & **Gurrib, I.** (2023). New era of artificial intelligence in education: Towards a sustainable multifaceted revolution. *Sustainability, 15*(16), 12451. https://doi.org/10.3390/su151612451

**Karpicke, J. D.,** & **Roediger III, H. L.** (2007). Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(4), 704. https://doi.org/10.1037/0278-7393.33.4.704

**Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ...** & **Kasneci, G.** (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

**Kaswan, K. S., Dhatterwal, J. S.,** & **Ojha, R. P.** (2024). AI in personalized learning. In *Advances in Technological Innovations in Higher Education* (pp. 103–117). CRC Press. https://doi.org/10.1201/9781003376699-9

**Kerfoot, B. P., Fu, Y., Baker, H., Connelly, D., Ritchey, M. L.,** & **Genega, E. M.** (2010). Online spaced education generates transfer and improves long-term retention of diagnostic skills: a randomized controlled trial. *Journal of the American College of Surgeons, 211*(3), 331–337. https://doi.org/10.1016/j.jamcollsurg.2010.04.023

**Koraishi, O.** (2023). Teaching English in the age of AI: Embracing ChatGPT to optimize EFL materials and assessment. *Language Education and Technology, 3*(1). https://langedutech.com/letjournal/index.php/let/article/view/48

**Lei, Q., Mason, R. A., Xin, Y. P., Davis, J. L., David, M.,** & **Lory, C.** (2020a). A meta–analysis of single–case research on mathematics word problem–solving interventions for English learners with learning disabilities and mathematics difficulties. *Learning Disabilities Research & Practice, 35*(4), 201–217. https://doi.org/10.1111/ldrp.12233

**Lei, Q.,** & **Xin, Y. P.** (2023). A synthesis of mathematical word problem-solving instructions for English learners with learning disabilities in mathematics. *Review of Education, 11*(2), e3396. https://doi.org/10.1002/rev3.3396

**Lei, Q., Xin, Y. P., Morita-Mullaney, T.,** & **Tzur, R.** (2020b). Instructional scaffolds in mathematics instruction for English learners with learning disabilities: An exploratory case study. *Learning Disabilities: A Contemporary Journal, 18*(1), 123–144. https://eric.ed.gov/?id=EJ1264271

**Lin, X.** (2023). Exploring the role of ChatGPT as a facilitator for motivating self-directed learning among adult learners. *Adult Learning,* 10451595231184928. https://doi.org/10.1177/10451595231184928

**Lin, X., Luterbach, K., Gregory, K. H.,** & **Sconyers, S. E.** (2024). A case study Investigating the utilization of ChatGPT in online discussions. *Online Learning, 28*(2), 1–23. https://doi.org/10.24059/olj.v28i2.4407

**Lin, X.,** & **Dai, Y.** (2022). An exploratory study of the effect of online learning readiness on self-regulated learning. *International Journal of Chinese Education, 11*(2). https://doi.org/10.1177/2212585X221111938

**Lin, X., Dai, Y.,** & **Shi, H.** (2022). Comparing the Sense of Online Classroom Community between Students from Urban and Rural Areas at a Chinese University. *Journal of Educators Online, 19*(3), n3. https://eric.ed.gov/?id=EJ1363837

**Liu, M., Ren, Y., Nyagoga, L. M., Stonier, F., Wu, Z.,** & **Yu, L.** (2023). Future of education in the era of generative artificial intelligence: Consensus among Chinese scholars on applications of ChatGPT in schools. *Future in Educational Research, 1*(1), 72–101. https://doi.org/10.1002/fer3.10

**Liyanage, M. L. A. P., Hirimuthugoda, U. J., Liyanage, N. L. T. N., Thammita, D. H. M. M. P., Wedanage, D. K. H. W., Kugathasan, A.,** & **Thelijjagoda, S.** (2022, October). AI solution to assist online education productivity via personalizing learning strategies and analyzing the student performance. In *2022 IEEE 13th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)* (pp. 0609–0615). IEEE. https://doi.org/10.1109/uemcon54665.2022.9965735

**Luan, L.** (2016). *Finding a basic interpretive unit through the human visual perception and cognition-A comparison between filmmakers and audiences. Bowling Green State University.* https://philpapers.org/rec/LUAFAB

**Luan, L.** (2021). *How readers process narrative information involving characters' decisions*. Bowling Green State University. https://www.proquest.com/docview/2627176621?pq-origsite=gscholar&fromopenview=true

**Luan, L., Huang, X., Zhu, S., Jiang, L., Chen, W.,** & **Ostadabbas, S.** (2023). An exploratory examination of online learning during and after the pandemic: Learning goal congruence in lecturing and research

activities. In *Rethinking Hybrid and Remote Work in Higher Education: Global Perspectives, Policies, and Practices after COVID-19* (pp.73–99). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-36632-1_5

**Luan, l., Lin, X., Dai, Y., Hu, S.,** & **Sun, Q.** (2024). Exploring the cognitive dynamics of artificial intelligence in the post-COVID-19 and learning 3.0 era: A case study of ChatGPT. *Asian Journal of Distance Education, 19*(2), 1–10. https://doi.org/10.5281/zenodo.10990472

**Madden, M. E., Baxter, M., Beauchamp, H., Bouchard, K., Habermas, D., Huff, M., ...** & **Plague, G.** (2013). Rethinking STEM education: An interdisciplinary STEAM curriculum. *Procedia Computer Science, 20*, 541–546. https://doi.org/10.1016/j.procs.2013.09.316

**Maghsudi, S., Lan, A., Xu, J.,** & **van Der Schaar, M.** (2021). Personalized education in the artificial intelligence era: what to expect next. *IEEE Signal Processing Magazine, 38*(3), 37–50. https://doi.org/10.1109/msp.2021.3055032

**Mann, H. B.,** & **Whitney, D. R.** (1947). On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, 50–60. https://doi.org/10.1214/aoms/1177730491

**Martin, S. M.** (2019). *Artificial intelligence, mixed reality, and the redefinition of the classroom*. Rowman & Littlefield.

**Michel-Villarreal, R., Vilalta-Perdomo, E., Salinas-Navarro, D. E., Thierry-Aguilera, R.,** & **Gerardou, F. S.** (2023). Challenges and opportunities of generative AI for higher education as explained by ChatGPT. *Education Sciences, 13*(9), 856. https://doi.org/10.3390/educsci13090856

**Mollick, E. R.,** & **Mollick, L.** (2024). Instructors as Innovators: a Future-focused Approach to New AI Learning Opportunities, With Prompts. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4802463

**Naseer, F., Khalid, M. U., Ayub, N., Rasool, A., Abbas, T.,** & **Afzal, M. W.** (2024). Automated assessment and feedback in higher education using generative AI. In *Transforming Education With Generative AI: Prompt Engineering and Synthetic Content Creation* (pp. 433–461). IGI Global. https://doi.org/10.4018/979-8-3693-1351-0.ch021

**Nazir, A.,** & **Wang, Z.** (2023). A comprehensive survey of ChatGPT: Advancements, applications, prospects, and challenges. *Meta-Radiology, 100022*. https://doi.org/10.1016/j.metrad.2023.100022

**Nepal, T. K.** (2024). Exploring the applications and challenges of chatGPT in research and academia: A comprehensive review. *West Science Interdisciplinary Studies, 2*(05), 1043–1050. https://doi.org/10.58812/wsis.v2i05.729

**Nicol, D. J.,** & **Macfarlane-Dick, D.** (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education, 31*(2), 199–218. https://doi.org/10.1080/03075070600572090

**Nielsen, F. Å.** (2011) A New ANEW: Evaluation of a Word List for Sentiment Analysis in Microblogs. *Proceedings of the ESWC2011 Workshop on "Making Sense of Microposts": Big Things Come in Small Packages*, Vol. 718, 93–98. https://www.scirp.org/reference/referencespapers?referenceid=3126488

**Nikolopoulou, K.** (2024). Generative artificial intelligence in higher education: Exploring ways of harnessing pedagogical practices with the assistance of ChatGPT. *International Journal of Changes in Education, 1*(2), 103–111. https://doi.org/10.47852/bonviewijce42022489

**Oleson, A., Hora, M.,** & **Benbow, R.** (2014). STEM: How a poorly defined acronym is shaping education and workforce development policy in the United States. https://eric.ed.gov/?id=ED556481

**Pedro, F., Subosa, M., Rivas, A.,** & **Valverde, P.** (2019). *Artificial intelligence in education: Challenges and opportunities for sustainable development*.

**Rane, N.** (2023). Enhancing mathematical capabilities through ChatGPT and similar generative artificial intelligence: Roles and challenges in solving mathematical problems. Available at SSRN 4603237. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4603237

**Rasul, T., Nair, S., Kalendra, D., Robin, M., de Oliveira Santini, F., Ladeira, W. J., ...** & **Heathcote, L.** (2023). The role of ChatGPT in higher education: Benefits, challenges, and future research directions. *Journal of Applied Learning and Teaching, 6*(1). https://doi.org/10.37074/jalt.2023.6.1.29

**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*. https://doi.org/10.1016/j.iotcps.2023.04.003

**Roediger III, H. L.,** & **Karpicke, J. D.** (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*(3), 249–255. https://doi.org/10.1111/j.1467-9280.2006.01693.x

**Sevnarayan, K.,** & **Potter, M. A.** (2024). Generative Artificial Intelligence in distance education: Transformations, challenges, and impact on academic integrity and student voice. *Journal of Applied Learning and Teaching, 7*(1). https://doi.org/10.37074/jalt.2024.7.1.41

**Slavin, R. E.** (1996). Research on cooperative learning and achievement: What we know, what we need to know. *Contemporary Educational Psychology, 21*(1), 43–69. https://doi.org/10.1006/ceps.1996.0004

**Stein, B., Haynes, A., Redding, M., Ennis, T.,** & **Cecil, M.** (2007). Assessing critical thinking in STEM and beyond. In *Innovations in E-Learning, Instruction Technology, Assessment, and Engineering Education* (pp. 79–82). Springer Netherlands. https://doi.org/10.1007/978-1-4020-6262-9_14

**STEM Education for the Future.** (2020, May). Retrieved from https://www.nsf.gov/edu/Materials/STEM%20Education%20for%20the%20Future%20-%202020%20Visioning%20Report.pdf

**Syahputra, M.** (2023). Exploring ChatGPT in language education: Friend or Foe for teachers? In *International Conference on Education, Science, Technology and Health (ICONESTH)* (pp. 1252–1269). https://eproceeding.bbg.ac.id/iconesth/article/view/221

**Tekin, C., Braun, J.,** & **van der Schaar, M.** (2015, April). eTutor: Online learning for personalized education. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 5545–5549). IEEE. https://doi.org/10.1109/icassp.2015.7179032

**Terry, W. S.** (2017). *Learning and memory: Basic principles, processes, and procedures*. Routledge. https://doi.org/10.4324/9781315622781

**Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R.,** & **Agyemang, B.** (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments, 10*(1), 15. https://doi.org/10.1186/s40561-023-00237-x

**Vahidy, J.** (2019). Enhancing STEM learning through technology. *Technology and the Curriculum: Summer 2019*. https://pressbooks.pub/techandcurr2019/chapter/enhancingstem/

**VanLehn, K.** (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist, 46*(4), 197–221. https://doi.org/10.1080/00461520.2011.611369

**Vashishth, T. K., Sharma, V., Sharma, K. K., Kumar, B., Panwar, R.,** & **Chaudhary, S.** (2024). AI-Driven Learning Analytics for Personalized Feedback and Assessment in Higher Education. In *Using Traditional Design Methods to Enhance AI-Driven Decision Making* (pp. 206–230). IGI Global. https://doi.org/10.4018/979-8-3693-0639-0.ch009

**Waghmare, C.** (2023). Security and ethical considerations when using ChatGPT. In *Unleashing The Power of ChatGPT: A Real World Business Applications* (pp. 111–132). Berkeley, CA: Apress. https://doi.org/10.1007/979-8-8688-0032-0_6

**Wang, Y., Tarakci, H.,** & **Prybutok, V.** (2023). Model comparison of regression, neural networks, and XGBoost as applied to the English Premier League transfer market. *International Journal of Sport Management and Marketing, 23*(6), 543–559. https://doi.org/10.1504/ijsmm.2023.133786

**Weimer, M.** (2013). *Learner-centered teaching: Five key changes to practice*. John Wiley & Sons. https://doi.org/10.2307/3211318

**Wu, T., He, S., Liu, J., Sun, S., Liu, K., Han, Q. L.,** & **Tang, Y.** (2023). A brief overview of ChatGPT: The history, status quo and potential future development. *IEEE/CAA Journal of Automatica Sinica, 10*(5), 1122–1136. https://doi.org/10.1109/jas.2023.123618

**Xie, X., Siau, K.,** & **Nah, F. F. H.** (2020). COVID-19 pandemic–online education in the new normal and the next normal. *Journal of Information Technology Case and Application Research, 22*(3), 175–187. https://doi.org/10.1080/15228053.2020.1824884

**Yesilyurt, Y. E.** (2023). AI-Enabled Assessment and Feedback Mechanisms for Language Learning: Transforming Pedagogy and Learner Experience. In *Transforming the Language Teaching Experience in the Age of AI* (pp. 25–43). IGI Global. https://doi.org/10.4018/978-1-6684-9893-4.ch002

**Zhong, L.** (2023). A systematic review of personalized learning in higher education: learning content structure, learning materials sequence, and learning readiness support. *Interactive Learning Environments, 31*(10), 7053–7073. https://doi.org/10.1080/10494820.2022.2061006

**TO CITE THIS ARTICLE:**
Luan, L., Lin, X., & Dai, Y. (2025). Bridging the Gap: ChatGPT's Role in Enhancing STEM Education. *Open Praxis*, *17*(1), pp. 108–128. DOI: https://doi.org/10.55982/openpraxis.17.1.685

**Submitted:** 21 March 2024
**Accepted:** 24 July 2024
**Published:** 15 April 2025