

Exploring the Feasibility and Efficacy of ChatGPT3 for Personalized Feedback in Teaching

Irum Naz and Rodney Robertson

Communications Department, University of Doha for Science and Technology, Qatar

irumzulfiqar@hotmail.com

rodney.d.robertson@gmail.com

<https://doi.org/10.34190/ejel.22.2.3345>

An open access article under [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License](#)

Abstract: This study explores the feasibility of using AI technology, specifically ChatGPT-3, to provide reliable, meaningful, and personalized feedback. Specifically, the study explores the benefits and limitations of using AI-based feedback in language learning; the pedagogical frameworks that underpin the effective use of AI-based feedback; the reliability of ChatGPT-3's feedback; and the potential implications of AI integration in language instruction. A review of existing literature identifies key themes and findings related to AI-based teaching practices. The study found that social cognitive theory (SCT) supports the potential use of AI chatbots in the learning process as AI can provide students with instant guidance and support that fosters personalized, independent learning experiences. Similarly, Krashen's second language acquisition theory (SLA) was found to support the hypothesis that AI use can enhance student learning by creating meaningful interaction in the target language wherein learners engage in genuine communication rather than focusing solely on linguistic form. To determine the reliability of AI-generated feedback, an analysis was performed on student writing. First, two rubrics were created by ChatGPT-3; AI then graded the papers, and the results were compared with human graded results using the same rubrics. The study concludes that e-Learning certainly has great potential; besides providing timely, personalized learning support, AI feedback can increase student motivation and foster learning independence. Not surprisingly, though, several caveats exist. It was found that ChatGPT-3 is prone to error and hallucination in providing student feedback, especially when presented with longer texts. To avoid this, rubrics must be carefully constructed, and teacher oversight is still very much required. This study will help educators transition to the new era of AI-assisted e-Learning by helping them make informed decisions about how to provide useful AI feedback that is underpinned by sound pedagogical principles.

Keywords: Artificial intelligence (AI), ChatGPT-3, Empowering teaching practices, Personalized feedback, Transformative implications

1. Introduction

The emergence of Large Language Model Artificial Intelligence (LLM AI) apps, particularly ChatGPT-3, has revolutionized pedagogical practices. While ChatGPT-3's advanced language generation and query-response capabilities offer great promise in enhancing language learning, its full potential and implications for education remain subjects of debate. This study aims to explore the feasibility of integrating AI technologies, specifically ChatGPT-3, into the teaching process to provide timely, effective, meaningful, and personalized feedback to students generally but language learners especially. The potential benefits of this kind of efficient feedback are indeed great when we consider that by 2025, eight million students will be studying internationally (Wang *et al.*, 2023). Not surprisingly, then, many educators and researchers have begun to explore the benefits and limitations of AI generated feedback. Wang *et al.* (2023), for example, have praised the timeliness of AI feedback while warning of inherent cultural biases in the AI evaluation process. Dai *et al.* (2023) emphasize the need to establish an effective feedback model by which to evaluate the efficacy of AI generated feedback. Researchers Buşe and Căbulea (2023) have serious reservations about AI's impact on creative thinking, human interaction, and technology dependence, while Cardon *et al.* (2023) argue that because AI-assisted writing is here to stay, instructors will have to greatly change how and what they teach.

This study hopes to add to the growing body of AI and education related literature by answering the following research questions:

- What are the benefits and limitations of using AI-based feedback in language learning and what pedagogical frameworks underpin the effective use of AI-based feedback?
- Is ChatGPT-3's feedback accurate and reliable enough to be effectively integrated into the teaching process to provide personalized feedback to language learners?
- What are the potential implications of AI integration in language instruction, and how can these findings contribute to the broader adoption of AI-based learning tools?

This research contributes to the understanding of how ChatGPT-3 can enhance language learning. The findings of this research can support educators and institutions in responsibly using AI to improve language proficiency and optimize the learning experience for second language learners. The research focuses on ChatGPT-3's potential for personalized feedback in language instruction, supported by theory and a review of relevant studies. The study relies on the researchers' experience grading student papers; AI-generated samples were graded by ChatGPT-3 and the quality of the AI's feedback was assessed based on a comparison of that feedback with human feedback.

2. Investigating AI's Potential Impact on Learning Feedback

According to Bandura (1977, pp.22), the process of learning is described as follows:

"Learning would be exceedingly laborious, not to mention hazardous, if people had to rely solely on the effects of their own actions to inform them what to do. Fortunately, most human behavior is learned observationally through modeling: from observing others, one forms an idea of how new behaviors are performed, and on later occasions, this coded information serves as a guide for action".

As Bandura highlights, modelling is essential to learning. The integration of AI chatbots, then, has the potential to revolutionize the learning landscape by offering learners the opportunity to observe, comprehend, and internalize feedback in novel ways. There have been numerous studies investigating performance feedback's impact on learning outcomes and its effectiveness in enhancing student performance. Feedback is a fundamental aspect of the learning process, playing a crucial role in shaping students' understanding, refining their skills, and fostering continuous improvement (Gray, Riegler, and Walsh, 2022). As Gray, Riegler, and Walsh (2022, pp. 16) point out, 74% of students agreed with the statements "It would have improved my performance if I had received more feedback on my work" and "When I received feedback on one piece of work, I used it when preparing a subsequent piece of work".

Previous studies have emphasized that timely and constructive feedback is a powerful tool for enhancing academic performance, promoting self-regulated learning, and nurturing a growth mindset among learners. In a 2011 experiment comparing one group of students who received immediate feedback with another group that received delayed feedback, Opitz, Ferdinand, and Mecklinger found that "the gain in performance was significantly larger for the group receiving immediate feedback as compared to the group receiving delayed feedback". Unfortunately, despite this obvious need for feedback, educators face challenges in delivering timely and adequate feedback to students.

Recently reported by Eric Bransteter (2022), four of the top causes of teacher burnout are: "long hours, large class sizes, additional responsibilities, [and having] too much on their plates." A series of recent studies has indicated that apart from the challenges posed by heavy workloads (Paris, 2022), large class sizes (Pisan et al., 2002), limited resources, and a lack of structured feedback systems, educators consistently confront distinctive hurdles in delivering timely and adequate feedback to students. For instance, the study conducted by Henderson et al. (2019) identified three key themes regarding feedback challenges: feedback practices, contextual constraints, and individual capacity. This study highlights feedback's complex interplay with practices, context, and individuals. Beyond known issues, it sheds light on "unique challenges", including producing meaningful personalized comments and addressing individual attitudes and capabilities.

Several authors have recognized that by harnessing the capabilities of AI, educators can help students navigate their learning journeys with greater efficacy and engagement (Biggam, 2010). Hwang and Chen (2023) identified six roles large language model AI can play in education: teacher/tutor, student/tutee, learning peer/partner, domain expert, administrator, learning tool. If this is true, as a tech savvy instructor would believe it is (with caveats), it seems AI tools such as ChatGPT-3 could function as personal virtual teachers/tutors and learning peers/partners with the ability to provide targeted, individualized, and instantaneous feedback. Educators have already begun to consider the application of AI technologies, particularly in the form of AI chatbots like ChatGPT-3, to provide personalized and meaningful feedback delivery (Mallow, 2023). What emerges now is the vital inquiry of how educators and experts perceive the merits and constraints of integrating AI-driven feedback into language learning.

The following section will explore the theoretical foundation of employing AI for language acquisition and its potential.

3. Theoretical Underpinnings: Converging SCT and Krashen's SLA Theory in AI-Facilitated Language Acquisition

Bandura's social cognitive theory (SCT) (1977) provides a robust theoretical foundation for exploring the cognitive intricacies of language acquisition, emphasizing self-efficacy and observational learning. In contrast, Krashen's second language acquisition (SLA) theory (1981) directs attention to the affective domain, highlighting the significance of meaningful communication and comprehensible input in language development. This analysis offers insight into how ChatGPT-3 can enhance comprehensive language acquisition by bridging both cognitive and affective aspects.

3.1 Reciprocal Interactions, Personalized Feedback, Cognitive Processes, and Learning from AI-Generated Feedback

In the context of AI-generated feedback, SCT provides valuable insights into how learners' cognitive processes, self-efficacy beliefs, and observational learning come into play when receiving and incorporating feedback from AI chatbots like ChatGPT.

SCT's principle of reciprocal determinism (Bandura, 1986) aligns with the concept of personalized and meaningful feedback provided by AI chatbots (Green, 2023). As learners engage with AI-generated feedback, their responses and subsequent learning behavior are influenced by the feedback itself, their pre-existing knowledge, and the learning environment. The AI chatbot, in turn, observes the learners' responses and generates subsequent feedback to better align with individual learning needs, creating a continuous feedback loop that fosters personalized learning experiences. As learners interact with AI chatbots to receive feedback, a dynamic and continuous feedback loop is established, wherein the AI chatbot observes the learners' responses and generates subsequent feedback tailored to their individual learning needs. This learning process engages learners in cognitive activities such as attention, perception, and memory to comprehend and internalize the AI-generated feedback. In this way, they benefit from the instant guidance and support that fosters personalized learning experiences. This reciprocal interaction between learners and AI chatbots, guided by SCT principles, facilitates the acquisition and integration of knowledge in a timely and tailored manner. Within the framework of SCT, the role of AI-generated feedback in the learning process can be explored through two distinct yet interconnected scenarios, each shedding light on the reciprocal interactions between learners and AI chatbots.

- **Teacher-Input Scenario:** In this situation, the teacher takes an active role in the feedback process. The teacher inputs the students' work into the AI chatbot, which then generates learning feedback based on the specific criteria provided by the teacher. The feedback is then delivered by the teacher to the students. Here, the AI chatbot acts as a tool that assists the teacher in providing personalized feedback to the learners.
- **Learner-Driven Scenario:** In this situation, learners themselves directly interact with the AI chatbot to receive, understand, and apply feedback. The AI chatbot is programmed to provide timely and tailored feedback to individual learners based on their responses and interactions. Learners take the initiative to seek feedback from the chatbot, which fosters independent learning and self-directed improvement.

These two stages represent different approaches to incorporating AI-generated feedback in the learning process. The first stage involves a more traditional setup where the teacher acts as an intermediary between the AI chatbot and the learners, facilitating the delivery and interpretation of feedback. The second stage, however, moves towards a more learner-driven model, where learners are trained to be actively engaged with the AI chatbot to receive personalized feedback, promoting self-regulated learning and autonomy.

The potential of ChatGPT in identifying and assisting learners in overcoming real-time challenges enhances motivation and engagement (Lin, 2023). The study conducted by Ali et al. (2023) investigates the impact of ChatGPT on the process of learning English. The results of the study indicated that ChatGPT serves as a source of motivation for learners, particularly in the enhancement of their reading and writing skills. However, when it comes to the development of listening and speaking skills, the respondents expressed relatively neutral attitudes. These findings collectively suggest that incorporating ChatGPT into English language teaching can be a motivational strategy with potential benefits for certain language skills.

Yet, it is important to acknowledge that this transformative technology is not without its hurdles. Scenario 2, the Learner-Driven scenario, requires careful consideration due to the risk of perpetuating biases within AI models, as highlighted by Ferrara (2023) when trained on biased datasets, potentially reinforcing harmful

stereotypes. Moreover, the limitations of AI in comprehending intricate or ambiguous input raise concerns about the quality of feedback it can offer. Chokwe (2015) observes that feedback provided to students in open and distance learning contexts is often insufficient, depriving them of valuable opportunities to learn from their mistakes. Burns (2010) emphasizes the potential loneliness of the distance learning experience, emphasizing the necessity for support and contact to ensure learners find value in the process.

3.2 Self-Efficacy and Response to Feedback

A central tenet of SCT is self-efficacy, which refers to individuals' beliefs in their ability to successfully execute specific tasks. In traditional educational settings, offering comprehensive feedback to a large cohort of students poses inherent challenges for instructors. Despite their well-intentioned efforts, instructors might inadvertently fail to bestow adequate emphasis on critical aspects of students' work, thus limiting the feedback's overall effectiveness. However, AI chatbots possess a unique capability to deliver personalized and equitable attention to each learner's performance. Through interactive exchanges with AI chatbots and the reception of feedback that acknowledges their exertions while imparting constructive guidance for refinement, learners' self-efficacy beliefs are bolstered. Vijayakumar, Höhn, and Schommer (2019) conducted a comprehensive study attesting to the research in this field and highlighting the potential of personalized and constructive feedback from 'conversational interfaces'. This feedback process, he asserts, has the power to significantly enhance learners' confidence in their abilities, fostering motivation to persist in their learning pursuits and embrace challenges. Consequently, the symbiotic relationship between learners and AI chatbots, grounded in SCT principles, nurtures an academic environment that fosters and empowers learners' self-efficacy beliefs, culminating in more effective, engaging, and transformative learning experiences.

Some may argue that such engagement is not possible without the emotional benefits of direct human interaction. However, support for the symbiotic nature of human-AI interaction can be deduced from studies on how video gamers interact emotionally with their video game avatars. A study (Hefner, Klimmt, and Vorderer, 2007, pp.46) has found that not only do game players identify with their virtual game personas and that this identification enhances enjoyment of the game but also that game players tend to strive to live up to the personal and professional expectations established by their game characters: "While it may be interesting to a given player to identify with the role of a corporate manager, our findings suggest that it is even more appealing to identify with a good manager, that is, to perform well within the role framework of the game". It is not inconceivable, therefore, that in the very near future, students will form a meaningful bond with their AI teacher/mentor/peers, and as Hefner, Klimmt, and Vorderer (2007) seem to suggest, this bond could help foster a healthy "striving to live up to" impulse in young people, thus enhancing their learning outcomes.

3.3 Observational Learning and Feedback Integration

One of ChatGPT's notable strengths is its capacity for self-learning, enabling a two-way learning process between users and the machine (Farrokhnia et al., 2023; Liu and Gibson, 2023). Research has provided evidence that when learners interact with AI chatbots and observe how feedback is generated and incorporated into their learning process, they essentially learn from the feedback generation process itself (Vijayakumar, Höhn, and Schommer, 2019b; Hancock et al., 2019). They observe how the AI chatbot analyzes their responses, identifies areas of improvement, and provides meaningful guidance for enhancement.

One study, Kostka and Toncelli (2023), highlights ChatGPT's capacity to revolutionize personalized support in the context of second (or subsequent) language development. A remarkable capability of AI chatbots is their ability to effortlessly generate correct model responses for learners, when guided by predefined rubrics and correction instructions. These model responses serve as exemplars of excellence, providing learners with clear benchmarks to strive towards in their own work. Learners can observe how the AI chatbot interprets their responses, identifies areas for improvement, and generates model answers that align with the prescribed rubrics. This process not only enhances learners' understanding of the feedback but also equips them with tangible examples of what constitutes a well-crafted response. As a result, learners can better comprehend the criteria used to evaluate their work and acquire a deeper understanding of the expected standards.

4. Navigating the Frontiers: Constraints and Challenges of ChatGPT-3

The integration of AI tools, such as ChatGPT, in education has the potential to enhance human intelligence (Carter and Nielsen 2017; Cotton, D. R., Cotton, P. A., and Shipway. 2023). Its contributions, however, coexist with counterarguments and limitations that warrant consideration (Koraishi, 2023). To effectively utilize ChatGPT in teaching and learning, it is essential to assess its capabilities as well as the constraints and challenges it offers. And, indeed, ChatGPT-3 exhibits some significant weaknesses.

Most relevant to this study, the quality of AI-generated responses must be carefully considered. As a large language model, ChatGPT-3 lacks a profound understanding of the words it processes, potentially resulting in ambiguous response outputs (Farrokhnia et al., 2023; Gao et al., 2023; Gupta, Raturi, and Venkateswarlu, 2023). There exists a considerable body of literature that claims that ChatGPT-3 struggles to evaluate the credibility of the data it was trained on, raising concerns about the quality and reliability of its responses (Farrokhnia et al., 2023; Lecler, Duron, and Soyer et al., 2023; Tlili et al., 2023). The potential for AI to produce errors or fabricate information, referred to as “hallucinations” (Randell and Coghlan, 2023), emphasizes the necessity for active teacher involvement to ensure responsible utilization of AI-generated materials. These quality concerns are the subject of the analysis in section 7 below.

Despite these limitations, however, the present study suggests that ChatGPT has the potential to serve as a valuable tool to foster student competence. Rather than replacing human intelligence, ChatGPT can enhance it when used under proper academic mentoring. As Kumar (2023) argues, it is important to recognize the limitations of AI tools and use them as teaching aids for students, not as replacement teachers. For instance, instructors can provide students with typical ChatGPT responses to assignments, highlighting the tool's shortcomings and offering recommendations for improvement.

5. AI vs. Human Teachers: Exploring the Educational Landscape

Human interaction, with its indispensable empathy and adaptability, plays a pivotal role in the learning journey, and AI should be seen as a supportive tool rather than a replacement for human educators. Chan and Tsi (2023) explore the potential of artificial intelligence (AI) in higher education, specifically its capacity to replace or assist human teachers. The study provides a comprehensive perspective on the future role of educators in the face of advancing AI technologies suggesting that although some believe AI may eventually replace teachers, most participants argue that human teachers possess unique qualities, such as critical thinking, creativity, and emotions, which make them irreplaceable. The study also emphasizes the importance of social-emotional competencies developed through human interactions, which AI technologies cannot currently replicate. Teachers need to understand how AI can work well with teachers and students while avoiding potential pitfalls, develop AI literacy, and address practical issues such as data protection, ethics, and privacy. The study reveals an interesting fact that students value and respect human teachers, even as AI becomes more prevalent in education.

This fact is yet reinforced by an AI invention proposal discussed in a recent study by Bakouan et al. (2018). Researchers created a chatbot model for responding to learners' concerns in online training. It uses a two-phase approach based on Dice similarity and domain-specific keywords. Notably, when a learner's question resembles a teacher's query, the chatbot asks for confirmation. If confirmed, it provides the relevant answer; if not, it redirects the query to a human tutor. The proposed AI invention highlights the importance of human intervention. The chatbot-human hybrid approach aims to enhance the learning experience. The article's process of transitioning from chatbot responses to human intervention in cases of complex queries highlights the essential role of human engagement in the learning process. It advocates for a hybrid teaching approach, where AI-driven chatbots support learners and collaborate with human educators to provide a more effective and personalized learning experience. This approach capitalizes on the strengths of both AI and human expertise, creating a foundation for enriched and adaptive learning (Bakouan et al., 2018).

6. Research Methodology: Assessing AI-Generated Student Feedback

This research study aims to provide an in-depth and comprehensive analysis of the utility and potential shortcomings of ChatGPT-3 in the context of student feedback, thereby contributing to the ongoing discourse surrounding AI in education.

6.1 Selection of Writing Samples

A diverse set of writing samples was carefully generated to represent a range of academic assignments. The study's sample comprised a 1000-word case-study paper (Student Paper 1) analyzing Tesla's marketing strategy, a shorter 100-word student reflection paragraph (Student Paper 2), and two 200-word summary paragraphs—one on the Impact of Climate Change and the other on *The Great Gatsby* (Student Paper 3 & 4). This selection of texts provided a relatively well-rounded basis for the research and ensured that the investigation covered distinct types of student-like work encountered in our teaching scenarios.

In addition to primary data collection through ChatGPT-3, our methodology incorporates an examination of the existing literature concerning AI-generated student feedback in education. This comprehensive literature review

serves to anchor our research in a robust understanding of the subject and entails a meticulous process of identifying relevant studies, assessing their quality, and conducting thorough content analysis. This approach ensures a holistic perspective on the effectiveness of ChatGPT-3 in delivering personalized feedback within the educational context.

6.2 Integration with ChatGPT-3

The selected writing samples were input into ChatGPT-3 and tasked with grading the mock assignments and generating feedback according to predefined criteria. These rubrics encompass a range of factors, including content, organization, clarity, and adherence to specific writing guidelines.

This step allowed for the examination how AI interacts with and assesses different types of student work while ensuring that the evaluation was aligned with the same standards applied to human grading. By using these criteria, the investigation aimed to maintain consistency and objectivity in the assessment process, thus facilitating a more accurate comparison of AI-generated feedback with human-generated feedback.

6.3 Content Analysis

Following data collection, a content analysis was undertaken to examine the feedback generated by ChatGPT-3. The analysis was focused on evaluating the accuracy and comprehensiveness of the AI-generated feedback, drawing a comparative assessment with feedback that human instructors might provide for identical student work. This comparative approach served as a benchmark to assess the effectiveness of AI in aligning feedback with educational goals. The content analysis involved detailed steps, including identification of key themes, assessment of feedback nuances, and categorization based on relevance to educational objectives. All pertinent data, encompassing mock student papers, instructor prompts, AI-generated feedback, and analytical notes, have been systematically documented to ensure transparency and facilitate research reproducibility.

7. Results and Interpretations: Effectiveness of ChatGPT3 for Personalized Feedback

This section offers a comprehensive examination of ChatGPT-3's feedback, including a thorough analysis of its assessment of mock student work and its potential implications within the educational context.

7.1 Novel Insights: Linking Krashen's Theories to ChatGPT-3's Language Acquisition Potential

An application of Krashen's language acquisition hypotheses to ChatGPT-3's response reveals the practical implications of using this AI technology. While prior research has explored ChatGPT-3's capabilities in language acquisition, our approach pioneers the interpretation that Krashen's theories lend crucial support to ChatGPT-3's vision and substantiate its utility in this context. This insight is particularly significant as it marks the first known attempt to bridge the gap between Krashen's theoretical framework and the real-world application of ChatGPT-3.

Krashen's Acquisition-Learning hypothesis explains the significance of meaningful interaction in the target language, wherein learners engage in genuine communication rather than focusing solely on linguistic form. This principle aligns with the capabilities of AI-generated feedback, exemplified by ChatGPT-3. Focusing on key terms 'personalized,' 'meaningful,' 'dynamic,' 'continuous,' and 'individuals' learning needs,' analysis of the data reveals that ChatGPT-3 is indeed capable of providing personalized and contextually relevant responses, potentially enabling learners to engage in language mirroring authentic communication as advocated by Krashen.

When a mock student text was input into ChatGPT-3 (see Figure 1 below), the AI system observed learners' responses and generated feedback tailored to their individual learning styles and needs. This established a continuous feedback loop, fostering a personalized and dynamic learning experience. As a result, this learning process actively engaged learners in cognitive activities such as attention, perception, and memory to comprehend and internalize the feedback.

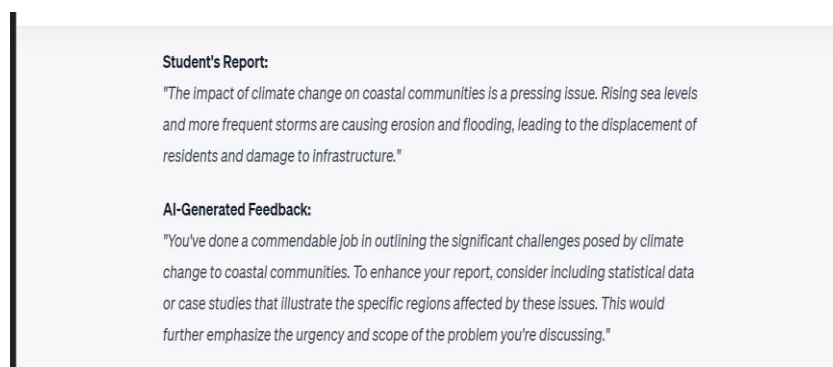


Figure 1: Student Paper 3: A Summary Paragraph on The Impact of Climate Change

The quality of AI-generated feedback can exhibit variations, yet the presence of key elements is consistently observed when well-engineered prompts are employed. As illustrated in Figure 1, the importance of keywords becomes evident. Firstly, the feedback encourages personalization by urging students to explore specific examples, thus tailoring the reports to the unique characteristics of 'distinct coastal regions.' The feedback places a strong emphasis on making the analysis meaningful, highlighting the significance of concrete evidence. The feedback promotes a dynamic and continuous approach by recommending the inclusion of various case studies, creating a flow of information that engages the learner. This dynamic and continuous engagement is facilitated by providing opportunities for ongoing feedback. The feedback ends with discussing individual's learning needs, encouraging exploration into how different coastal areas are uniquely affected. By individually addressing learning needs, students are more likely to engage effectively with the material, attain a deeper understanding of the subject matter, and ultimately experience improved learning outcomes.

Krashen's Acquisition-Learning hypothesis highlights the value of meaningful interaction in the target language, wherein learners engage in authentic communication rather than solely focusing on linguistic form. AI-generated feedback, exemplified by ChatGPT-3, aligns with this principle. The findings demonstrate that the AI bot is adept at delivering personalized and contextually relevant responses, potentially enabling learners to engage in language use that mirrors the essence of genuine communication, as advocated by Krashen. The feedback in Figure 1 is distinctly personalized and contextually relevant, directly addressing the content of the student's report. It not only acknowledges the student's effort but also provides specific and constructive guidance on how to enhance the work by suggesting the inclusion of statistical data or case studies. The feedback is intricately tailored to the topic of climate change and coastal communities, reflecting a clear understanding of the context and the student's work.

This alignment with the student's individual needs and the contextual relevance of the feedback is striking. According to Krashen's Input hypothesis, a fundamental principle of language acquisition posits that learners progress by encountering 'Comprehensible Input' slightly more complex than their current level of competence. This theory emphasizes the importance of exposure to 'i + 1,' which represents the next linguistic stage in the learner's development and promotes language growth. In this very context, ChatGPT-3's AI-generated feedback emerges as a transformative force. By dynamically generating tailored input that aligns with the learner's proficiency level, it effectively offers a continuum of comprehensible language exposure.

The response provided in Figure 2 below acknowledges Alex's solid grasp of the green light's symbolism and suggests a deeper exploration of the color green in literature. Also, it encourages the student to delve into the psychology of colors, contributing to a dynamic and continuous learning process.

Report Excerpt (Alex's Submission)

Excerpt from Alex's report:
"In 'The Great Gatsby,' the green light symbolizes hope and aspiration. This recurring motif appears throughout the novel, representing Gatsby's unfulfilled dreams."

ChatGPT3 Feedback

AI-generated feedback:
"Great job, Alex! Your analysis of the green light's symbolism is on point. Consider exploring the nuances of the color green in literature. Delve deeper into the psychology of colors and how it connects to Gatsby's character. Additionally, you could incorporate parallels with other literary works that use similar motifs. Keep up the excellent work!"

Figure 2: Student Paper 4: A Summary Paragraph on The Great Gatsby

In addition, the recommendation to draw parallels with other literary works is not merely a casual suggestion; it plays a pivotal role in enhancing Alex's ability to establish meaningful connections across texts. This, in turn, fosters a richer and more comprehensive understanding of symbolism in literature, reflecting the feedback's commitment to promoting individualized learning tailored to Alex's specific needs. It effectively encourages him to expand his analytical capabilities in a contextual and meaningful manner.

The discussion thus far, encompassing the principles of SCT (see Sections 3.1 – 3.3) and two SLA acquisition hypotheses, leads to a consideration of SLA's Affective Filter Hypothesis. Krashen's hypothesis posits that various emotional factors contribute, in a non-causal manner, to second language acquisition, with learners exhibiting high motivation, self-confidence, a positive self-image, and lower anxiety tending to excel. While not directly supported by our study or existing literature, it's worth contemplating how AI may potentially reduce anxiety levels among learners. Interactions with AI systems often occur in an environment devoid of social pressures and judgments. Learners tend to find seeking guidance from AI more comfortable, enabling them to engage more openly without the fear of making mistakes or being judged for them. Moreover, the timely and personalized nature of AI-generated feedback caters to individual needs, potentially enhancing learners' self-confidence by providing constructive insights for improvement.

Despite these potential strengths, however, an analysis of the mock student papers in the next sections reveals several weaknesses of AI as feedback provider.

7.2 Student Paper 1, Prompt 1

Figure 3 outlines the instructor’s prompts, ChatGPT’s responses, and analysis notes of that response.

Instructor Prompts	ChatGPT Responses
	NOTE: Responses generated by same prompts on different ChatGPT accounts on different days
Based on the following Case Study show evidence that the student has a good understanding of the key concepts and theories covered in the course, is able to apply his/her knowledge to real-world business scenarios or case studies, and can analyze complex information and think critically about business problems and situations.	Response 1: Great work on analyzing Tesla's market expansion strategy! It is clear that you have a strong understanding of the key concepts and theories covered in the course and are able to apply them to real-world business scenarios. You demonstrate strong critical thinking and analysis skills and are able to communicate your findings effectively. Your use of evidence and research, attention to detail, and creativity and innovation are all impressive. Your teamwork and collaboration, time management and organization, customer focus and market research, and project management and accountability skills are also strong. Finally, your attention to ethics and social responsibility is commendable. Keep up the good work! Response 2: The case study provided by the student demonstrates several indications of a good understanding of key concepts and theories covered in the course, application of knowledge to real-world business scenarios or case studies, and critical thinking skills.
Analysis	
Vague. No examples.	Firstly, the student demonstrates an understanding of the concept of marketing mix, which is a fundamental marketing concept. In the case study, the student analyzes how the company's marketing mix affects its sales and revenue. This indicates a good understanding of the marketing mix concept and the ability to apply it to a real-world business scenario.
Merely regurgitates keywords from prompt.	Secondly, the student uses data analysis to identify the company's strengths and weaknesses. This demonstrates the ability to analyze complex information and think critically about business problems and situations.
Fabrications, hallucinations	Thirdly, the student suggests several recommendations to improve the company's sales and revenue, based on the analysis of the company's strengths and weaknesses. This indicates the ability to apply knowledge to real-world business scenarios or case studies and to think critically about business problems and situations.
The submission only claims to have performed data analysis; it contains no real analysis.	

Figure 3: Instructor prompts, AI feedback, and Analysis Notes for Student Paper 1: Tesla’s Marketing Strategy

As can be seen from the colour-coding in Figure 3, understanding the key concepts, applying those concepts, and analyzing complex information were the three criteria by which student paper 1 was assessed. The criteria keywords used in ChatGPT-3's responses are color-coded. At first glance, the AI seemed to have done at least a fair job of identifying where the student succeeded in satisfying the assignment's criteria. Through the repetition of the keywords, ChatGPT-3 demonstrated an ability to take the provided criteria and apply it to the student work (See Response 1 in Figure3).

A careful examination of what the text says, however, quickly reveals that the feedback's content lacks precision and coherence. A significant portion of the content appears to be devoid of substantive meaning, as it predominantly consists of a mere reiteration of keywords derived from the initial prompt. Notably, the AI's initial response to the prompt exhibits a particularly pronounced deficiency in terms of content quality; it provides ample praise for the student work, but no evidence of quality. Vagueness is substituted for depth of analysis. In this example, the AI responds with, "It is clear you have a strong understanding of the key concepts and theories", without specifying the exact location where this understanding is demonstrated. It gets worse. The AI also responded: "In the case study, the student analyzes how the company's marketing mix affects its sales and revenue. This indicates a good understanding of the marketing mix concept and the ability to apply it to a real-world business scenario." In fact, Paper 1 contains no mention at all of "marketing mix", only the word "marketing", and certainly no analysis of the concept. The request for a second response based on the same prompt resulted in essentially more of the same, as demonstrated in Figure 3, Response 2.

The AI exhibited several behaviors during the analysis. Firstly, it appeared to be merely searching for keywords from the prompts in the student paper, words such as "analysis", "marketing", etc. Secondly, if it found a word, it responded with a "well done". Worse than this, however, it seemed that when the AI could not find a suitable keyword or phrase, it hallucinated a response, as if it wanted to please the user, or was determined to satisfy the prompt's demands regardless of the veracity of its responses. Analysis is a very complex cognitive process, and the AI could not identify whether the student was demonstrating that skill in the paper. Paper 1 itself is not very sophisticated, and neither are many real student papers; if AI is going to become a trusted feedback provider, it needs to be able to tell the difference between hallucination or empty keyword regurgitation and true analysis.

7.3 Student Paper 1, Prompt 2

Continuous feedback loop theory was applied to create a second prompt to improve results. See Figure 4 below.

Instructor Prompts	ChatGPT Responses
Give categorical examples of the above given skills from student's work.	Sure, here are some examples of the student demonstrating their understanding of key concepts, applying their knowledge to real-world scenarios, and analyzing complex information in the case study:
Analysis	Understanding of key concepts:
False response; submission only mentions supply chain; no mention of other concepts.	- The student accurately defined and explained terms such as "supply chain," "vendor-managed inventory," and "just-in-time" inventory management throughout the case study. - The student demonstrated a clear understanding of the different factors that affect demand and supply, including seasonality, market trends, and customer behavior.
Keyword regurgitation; fabrication	Application of knowledge to real-world scenarios:
Hallucination; the submission is not about a toy company	- The student applied their knowledge of inventory management to propose solutions to the challenges faced by the toy company in the case study, such as implementing a vendor-managed inventory system. - The student used their understanding of supply chain management to suggest ways that the toy company could streamline their processes and reduce costs, such as by partnering with a logistics company to optimize their shipping routes.
	Analysis of complex information:
	- The student analyzed the toy company's financial statements to identify areas of inefficiency and suggest improvements. - The student evaluated the pros and cons of various solutions to the toy company's inventory and supply chain challenges, considering factors such as cost, feasibility, and potential impact on the company's operations.

Figure 4: Second Instructor Prompts, AI Feedback, and Analysis Notes for Student Paper 1: Tesla's Marketing Strategy

By requesting that ChatGPT-3 provide categorical examples of the skills claimed to have been demonstrated in Prompt 1, an effort was made to enhance the AI's initial vague and inaccurate assessment. Unfortunately, the results were worse; there was more hallucination and more keyword regurgitation. For example, the terms "vendor-managed inventory" and "just-in-time inventory management" were not found in Paper 1 at all. ChatGPT is a sophisticated AI language model, and it appears to have followed its programming by searching the

internet for relevant concepts it believed would meet the prompt, delivering what it perceived as the desired response. The only concept from ChatGPT's responses in Figure 4 that was also mentioned in the actual student paper was the concept of "supply chain". It also invented the idea of "toy company". Of course, the paper was not about a toy company; it was about Tesla's marketing strategy. Once more, the AI provided examples of marketing strategies and concepts that it believed the prompts were seeking.

7.4 Student Paper 2, Prompt 1

Figure 5 outlines the instructor's prompts, ChatGPT's responses, and analysis notes of that response for student Paper 2.

Instructor Prompts	ChatGPT Responses
Please evaluate the given paragraph based on the following criteria: 1. Identification of the main idea, key supporting details, and use of effective evidence to support the argument. 2. Clarity, conciseness, and coherence in transitioning between paragraphs and ideas. 3. Identification of any areas that could be improved, and suggestions for future steps to improve the writing. Analysis Quite good general writing advice. Specific weaknesses are identified AI demonstrates a good understanding of paragraph structure requirements and can identify	Regarding the first criterion, the paragraph does touch upon several potential main ideas, including the importance of the survey for the author's future career aspirations, financial goals, desire to impress others, and personal growth. However, the paragraph lacks key supporting details or effective evidence to support these ideas. For instance, the author mentions wanting to improve their reading and writing skills, but does not explain how the survey would help them achieve this. In terms of the second criterion, the paragraph is also unclear and lacks conciseness. The author jumps from one idea to another without effective transitions or clear connections between them. Additionally, the use of language is sometimes awkward, such as the use of the phrase "to have a fun job" instead of "to have a fun job." Finally, regarding the third criterion, there are several areas in which the writing could be improved. The author could benefit from focusing on one or two main ideas and providing more detailed supporting evidence for each. Additionally, the author could work on improving the clarity and coherence of their writing, perhaps by using more transitional phrases and clear topic sentences. In terms of next steps, the author could consider revising the paragraph with a more focused and organized approach, while also working on improving their writing skills through practice and feedback.

Figure 5: Instructor Prompts, AI Feedback, and Analysis Notes for Student Paper 2: Reflection Paragraph

Paper 2 was a much simpler mock student work and seemed to be better suited to the type of feedback ChatGPT-3 is capable of. The assignment merely required the student to express their opinion about whether a survey they had completed was useful to them. They were to write in a paragraph form. As a short personal reflection, the paragraph was not as content and concept heavy as the longer Paper 1. As can be seen from Figure 5, the criteria were comprised of two standard paragraph assessment items: (1) identification of the main idea, key supporting details, and use of effective evidence to support the argument and (2) clarity, conciseness, and coherence in transitioning. A third prompt asked for suggestions for future steps to improve the writing.

This time, ChatGPT's feedback did more than simply hit upon keywords, although it did do that, too; see colour-coding in Figure 4. The AI demonstrated a good understanding of paragraph structure requirements and could identify those elements quite accurately in the student work: "Regarding the first criterion, the paragraph does touch upon several potential main ideas, including the importance of the survey for the author's future career aspirations, financial goals, desire to impress others, and personal growth" (Figure 4). More impressive still, the AI correctly pointed out that "the paragraph lacks key supporting details or effective evidence to support these ideas", and provided examples: "For instance, the author mentions wanting to improve their reading and writing skills but does not explain how the survey would help them achieve this" (Figure 4). As instructors, there is consensus that this constitutes an accurate assessment of the student's work.

Regarding the second criteria, the AI's assessment was more or less accurate as well, but slightly more critical than would be acceptable. Paper 2 lacked some transitioning and coherence devices and jumped from one idea to the next, somewhat. But the paragraph was certainly not "unclear" (Figure 4). The language, too, was not particularly awkward, the AI having mistaken incorrectness for casual tone in the phrase "to have a fun job" (Figure 4).

The ultimate criterion, which pertains to advice for future improvements in writing, was deemed fitting and closely aligned with the guidance instructors typically offer to students: "The author could benefit from focusing on one or two main ideas and providing more detailed supporting evidence for each" (Figure 4). This is very typical advice teachers give to student writing, so it is possible that ChatGPT merely assembled it from its internet readings. Nevertheless, as applied to Paper 2, the advice is apt and accurate.

7.5 Student Paper 2, Rubric Creation, and Paragraph Grading

The first attempt at using ChatGPT-3 for rubric generation for Paper 2 was not successful, resulting in a grade that a human instructor would certainly not have given the paper. See Figure 6 below.

Instructor Prompts	ChatGPT Responses		
	Criteria	Score	Description
Please create a rubric for evaluating paragraphs, including the following criteria with a weight of 2 scores for each item: Spelling Errors, Grammatical Errors, Run-on Sentences, Sentence Fragments, Wordiness, Incorrect Word Choice, and Lack of Clarity. Ensure that the rubric is well-structured and clearly defines each of these categories. Evaluate the following paragraph based on the above given rubric and produce a tabulated feedback, "..." Analysis Identification of errors is fair AI makes 3 errors Total score of 6/14 (43%) is much too low. Student manages to convey her ideas quite well despite clumsiness of language; a more accurate score would be 65 or 70%.	Spelling Errors	1	"havea" should be "have a"
	Grammatical Errors	1	"as soon I finish the college" should be "as soon as I finish college"
	Run-on Sentences	0	The paragraph contains multiple run-on sentences. For example, the first sentence should be separated into at least two sentences. FAIR. AWKWARD BUT NOT QUITE A RUN-ON
	Sentence Fragments	1	The sentence "To improve your learning" is a sentence fragment, as it lacks a subject and verb. WRONG. THE ORIGINAL PHRASE ENDS WITH A COMMA
	Wordiness	1	The sentence "Also, you need to think about the writing that you have written" could be simplified to "Think about your writing."
	Incorrect Word Choice	1	The sentence "To improve your learning, you have to understand what are writing about" should be "To improve your learning, you have to understand what you are writing about."
	Lack of Clarity	1	The paragraph lacks clear organization and structure, making it difficult to follow. Additionally, some sentences are unclear or do not make complete sense. NOT TRUE

Figure 6: Instructor Prompts for Rubric Creation and Assessment, AI Feedback, and Analysis Notes for Student Paper 2: Reflection Paragraph

As demonstrated in Figure 6, the AI's identification of some errors was fair. But it made several serious errors in assessment. For example, a comma splice was misidentified as a sentence fragment. More egregiously, under "lack of clarity", the AI claimed the paragraph lacks clear organization and structure. Although some transitions were lacking, Paper 2 was certainly not "difficult to follow" (Figure 6). Additionally, the claim that "some sentences are unclear or do not make complete sense" is blatantly false (Figure 6).

On the two-point scale, ChatGPT-3 scored Paper 2 a 6/14 (43%), which was not at all appropriate. Most instructors would probably give this paper around a 65-70%. The AI seemed to be incapable of ignoring the minor mistakes when appropriate and giving points and credit for overall understandability and readability. The paragraph was quite readable; it made logical sense and contained several good ideas. As discussed, the paper lacked some development and transitioning, but all the ideas were clearly expressed.

In the second iteration of rubric generation and evaluation, a strategy was implemented involving the utilization of more specific prompt questions directed at the AI, as illustrated in Figure 5. This adjustment yielded significantly improved grading accuracy for Paper 2.

Instructor Prompts	ChatGPT Responses		
	Criteria	Score	Description
Design a tabulated rubric including the following: 1. Identify the criteria that will be used to evaluate the performance. 2. Define each criterion and describe what it means. 3. Determine the scale of evaluation, such as A to F or 1 to 5. 4. Create verbal descriptors for each level of the evaluation scale. 5. Tabulate the rubric, listing each criterion, its description, and the corresponding verbal descriptor for each level of the evaluation scale. Evaluate the following paragraph based on the above given rubric. Write the feedback in a table format. Analysis Good, useful feedback, but lacking examples. More accurate evaluation: 63% is much closer to the grade I would have assigned; a bit low. A broader range of potential marks allowed for more flexible evaluation. Again, AI does not seem to take readability into account.	Spelling Errors	4	The presence of misspelled words or typos in the text
	Grammatical Errors	3	Errors in sentence structure or grammar
	Run-on Sentences	3	Sentences that are too long or improperly joined, making the text difficult to follow
	Sentence Fragments	3	Incomplete sentences or phrases that do not form a complete thought
	Wordiness	3	The use of excessive or unnecessary words or phrases, making the text unnecessarily long or difficult to read
	Incorrect Word Choice	3	The use of incorrect or inappropriate words or phrases, leading to confusion or ambiguity
	Lack of Clarity	3	The overall lack of clarity or coherence in the paragraph, making it difficult to follow or understand

Figure 7: Instructor Prompts for Second Rubric Creation and Assessment, AI Feedback, and Analysis Notes for Student Paper 2: Reflection Paragraph

As can be seen in Figure 7, the descriptive assessment was vague and lacked examples, and thus, although more or less accurate, not particularly useful. For example, the AI claimed Paper 2 was difficult to follow, which, as discussed, was not the case.

The grade assessment of 63%, though, was more accurate and useful. The two-point scale used in the first rubric was obviously too narrow, resulting in something approximating a binary right or wrong assessment of items that may not be perfect, but that do not merit a score of zero. The broader range of potential marks of the second rubric's five-point scale allowed for more flexible evaluation. Again, the AI does not seem to take readability into account, which is an important drawback because students are not robots. They communicate in many ways—on paper as well as off paper. Humans instructors can take this into account when assessing student work.

8. Conclusion and Recommendations

In the course of this study, an examination has been conducted regarding the benefits and limitations of using AI-based feedback, the pedagogical frameworks underpinning effective utilization of such feedback, and the feedback's accuracy and reliability for integration into the teaching process. Furthermore, the potential implications of AI integration in language instruction and its contribution to the broader adoption of AI-based learning tools has been explored. The following discussion further elucidates how these critical inquiries have been addressed and clarified by the findings.

It is evident that teachers who are eager to harness the substantial potential of AI to deliver timely and precise student feedback, especially when they may have limited time for such tasks, are advised to exercise caution. This study has revealed several implications in this regard. This section will discuss both the potential benefits and drawbacks by revisiting and addressing study's initial research questions.

Perhaps most obviously, the absence of the human touch and personalized approach may hinder the complete fulfillment of learners' specific needs. As AI's understanding of human behavior and needs becomes more sophisticated, it may be able to accomplish many of the goals performed by human mentors, teachers, and peers. Currently, however, AI's limitations in capturing the complexities of language learning, particularly in pragmatics and sociolinguistics, call for a cautious integration of AI-generated feedback in language instruction.

It is essential to acknowledge that while AI-generated feedback may offer general improvements, it may fall short in addressing the individual challenges and requirements of diverse learners. The limitations of AI in understanding complex or ambiguous input raise concerns about the quality of feedback it may provide.

ChatGPT-3 seems competent at assessing simple, clearly focused language criteria. It did an adequate job analyzing the student's 100-word paragraph for main idea, support, coherence, and conciseness. However, the AI seems to have difficulty analyzing students' critical thinking skills. The lower levels of Bloom's taxonomy – remembering and understanding – seem to be within its realm of competence. See Figure 8 below. It is, after all, a rather mechanical task for the AI to locate evidence of student recall and interpretation of main ideas and learned material. Higher, more abstract, and more human skills may be beyond AI's powers at this stage. In our experiment, for example, ChatGPT-3 had difficulty evaluating student application, analysis, and evaluation of marketing concepts.

Remembering	Understanding	Applying	Analyzing	Evaluating	Creating
Demonstrate memory of learned material through recall.	Demonstrate understanding of learned material by organizing, comparing, translating, interpreting and stating main ideas.	Solve problems in new situations by applying learned knowledge and techniques.	Examine and separate information by identifying motive or cause. Make inferences and identify evidence to support.	Present and defend options by making judgments about information, idea validity, quality based on learned information.	Compiled information in a new and effective way through the development of a pattern or solution that differs from those learned.
Verbs: Choose, Define, Describe, Find, How, Label, List, Locate, Match, Name, Locate, Omit, Recall, Relate, Select, Show, Tell, When, Where, Which, Who, Why	Verbs: Classify, Compare, Contrast, Demonstrate, Explain, Exemplify, Extend, Infer, Interpret, Infer, Paraphrase, Relate, Summarize, Translate	Verbs: Apply, Build, Carry out, Choose, Construct, Develop, Execute, Experiment with, Identify, Interview, Implement, Make use of, Model, Organize, Plan, Select, Solve, Use, Utilize	Verbs: Analyze, Assume, Categorize, Classify, Compare, Conclude, Contrast, Discover, Dissect, Distinguish, Divide, Deconstruct, Examine, Function, Inspect, Integrate, Simplify, Survey, Take part in, Test, Outline, Deconstruct	Verbs: Agree, Appraise, Argue, Assess, Award, Choose, Compare, Critique, Decide, Deduct, Defend, Determine, Disprove, Estimate, Evaluate, Explain, Influence, Judge, Justify, Measure, Perceive, Prioritize, Prove, Rate, Recommend, Support, Value	Verbs: Adapt, Build, Change, Choose, Combine, Compile, Compose, Construct, Create, Delete, Design, Develop, Discuss, Elaborate, Estimate, Formulate, Improve, Invent, Make Up, Maximize, Minimize, Modify, Originate, Plan, Predict, Propose, Solve, Suppose, Theorize

ChatGPT3's zone of competence regarding student feedback

Figure 8: Revised Bloom's Taxonomy - Adapted from Anderson and Krathwohl (2001)

AI feedback may be useful for assessing assignments in courses such as marketing, engineering, or health sciences, only when content knowledge is being tested. Application of knowledge and creative thinking should be assessed by the instructor.

Rubric generation and grading require careful attention. The more precise the criteria, the more accurate the evaluation. ChatGPT-3 adhered to the Rubrics too rigidly, marking grammar and style too harshly; it seemed unable to account for human comprehensibility despite minor errors.

Although beyond the scope of this study, it is worth considering that the vagueness of ChatGPT-3's responses has implications for teaching students how to use AI effectively in their own research. In short, technical students relying on ChatGPT-3 to write reports are going to be disappointed with the vague content it produces, of little practical or theoretical use.

The utilization of well-suited prompts holds the potential to yield outcomes that closely correspond to users' expectations. This notion is substantiated by the observation that well-tailored prompts contribute to feedback outputs that exhibit a greater resonance with the intended context and purpose of the feedback. In essence, the careful engineering of prompts facilitates a more nuanced and contextually relevant interaction with AI systems, thereby enhancing the quality and relevance of the generated feedback.

9. Further Research

To further validate the generalizability of this study, future studies should include a broader range of educational contexts, such as humanities papers, lab reports, or creative writing. TVET and job training are other areas for potential use of AI generated feedback and may prove suitable domains for evaluation of the lower end of Bloom's Taxonomy.

Future iterations of ChatGPT and other large language model AI will, of course, produce much more accurate student feedback. The researchers involved in this study anticipate that forthcoming analytical investigations, like this one, will unveil the potential for AI to be employed with assurance by both instructors and students, ensuring the provision of meaningful and timely feedback. The potential is limitless, and educators must make good use of it. In fact, educational AI software could be designed specifically to provide student feedback. It could be pre-programmed with learning theories such as the ones discussed here (Social Cognitive Theory and Second Language Acquisition Theory) and applied directly to student work. Feedback generated in this way has the potential to greatly enhance autonomous student learning. In the meantime, educational scholars must continue to monitor the rapid progress of LLM AI such as ChatGPT for veracity and accuracy by performing critical reviews on the impact of AI on student learning and analysis studies on AI's feedback performance.

References

- Ali, J. K. M., Shamsan, M. a. A., Hezam, T. A., and Mohammed, A. a. Q., 2023. Impact of ChatGPT on Learning Motivation. *Journal of English Studies in Arabia Felix*, 2(1), pp 41–49. <https://doi.org/10.56540/jesaf.v2i1.51>
- Anderson, L. W. and Krathwohl, D. R., 2001. *A taxonomy for learning, teaching, and assessing, Abridged Edition*. Boston: Allyn and Bacon.
- Bandura, A., 1986. *Social Foundations of Thought and Action: A Social Cognitive Theory*. London: Prentice-Hall.
- Biggam, J., 2010. Reducing staff workload and improving student summative and formative feedback through automation: squaring the circle. *International Journal of Teaching and Case Studies*, 2(3/4), 276. <https://doi.org/10.1504/ijtc.2010.033322>
- Bransteter, E. 2022. What are the most common causes of teacher burnout?. *iAspire Education*. [online] Available at <<https://iaspireapp.com/resources/what-are-the-most-common-causes-of-teacher-burnout#:~:text=Whether%20it%20be%20from%20challenging,lack%20of%20connection%20and%20support.>> [Accessed 13 September 2023].
- Burns, M., 2010. Not Too Distant: A Survey of Strategies for Teacher Support in Distance Education Programs. *The Turkish Online Journal of Distance Education*, 11(2), pp 108–117. <https://doi.org/10.17718/tojde.88297>
- Bușe, O. and Căbulea, M. (2023) 'Artificial intelligence – an ally or a foe of foreign language teaching?', *Revista Academiei Forțelor Terestre*, 28(4), pp. 277–282. <https://doi.org/10.2478/raft-2023-0032>.
- Cardon, P.W. et al. (2023) 'The Challenges and Opportunities of AI-Assisted Writing: Developing AI Literacy for the AI Age,' *Business and Professional Communication Quarterly*, 86(3), pp. 257–295. <https://doi.org/10.1177/23294906231176517>.
- Carter, S., and Nielsen, M., 2017. Using artificial intelligence to augment human intelligence. *Distill*, 2(12). <https://doi.org/10.23915/distill.00009>
- Chan, C. K. Y., and Tsi, L. H. Y., 2023. The AI revolution in Education: Will AI replace or assist teachers in higher education? *arXiv*. Cornell University. <https://doi.org/10.48550/arxiv.2305.01185>
- Chokwe, J. M., 2015. Students' and tutors' perceptions of feedback on academic essays in an open and distance learning context. *Open Praxis*, 7(1), p 39. <https://doi.org/10.5944/openpraxis.7.1.154>
- Cotton, D. R., Cotton, P. A., and Shipway, J., 2023. Chatting and cheating. Ensuring academic integrity in the era of ChatGPT. *EdArXiv*. [online] Available at <<https://edarxiv.org/mrz8h/>> [Accessed 25 June 2023].

- Dai, W. et al. (2023) 'Can Large Language Models Provide Feedback to Students? A Case Study on ChatGPT,' *Centre for Learning Analytics at Monash, Monash University* [Preprint]. <https://doi.org/10.35542/osf.io/hcgzj>.
- Farrokhnia, M., Banihashem, S. K., Noroozi, O., and Wals, A., 2023. A SWOT analysis of ChatGPT: Implications for educational practice and research. *Innovations in Education and Teaching International*. <https://doi.org/10.1080/14703297.2023.2195846>
- Ferrara, E., 2023. Should ChatGPT be Biased? Challenges and Risks of Bias in Large Language Models. *arXiv*. Cornell University. <https://doi.org/10.48550/arxiv.2304.03738>
- Gao, J., Zhao, H., Yu, C., and Xu, R., 2023. Exploring the feasibility of ChatGPT for event extraction. *arXiv*. <https://doi.org/10.48550/arXiv.2303.03836>
- Gray, K., Riegler, R., and Walsh, M. 2022, January 29. Students' feedback experiences and expectations pre- and post-university entry. *SN Social Sciences*, 2(2). <https://doi.org/10.1007/s43545-022-00313-y>
- Gupta, P. K., Raturi, S., and Venkateswarlu, P., 2023. Chatgpt for designing course outlines: A boon or bane to modern technology. <https://doi.org/10.2139/ssrn.4386113>
- Hancock, B., Bordes, A., Mazaré, P., and Weston, J., 2019. Learning from Dialogue after Deployment: Feed Yourself, Chatbot! *ACL Anthology, 57th Annual Meeting of the Association for Computational Linguistics*. <https://doi.org/10.18653/v1/p19-1358>
- Hefner, D., Klimmt, C., and Vorderer, P., 2007. Identification with the Player Character as Determinant of Video Game Enjoyment. *Lecture Notes in Computer Science*, pp 39–48. https://doi.org/10.1007/978-3-540-74873-1_6
- Hwang, G.-J., and Chen, N.S., 2023. Editorial Position Paper: Exploring the Potential of Generative Artificial Intelligence in Education: Applications, Challenges, and Future Research Directions. *Educational Technology & Society*, 26(2). [https://doi.org/10.30191/ETS.202304_26\(2\).0014](https://doi.org/10.30191/ETS.202304_26(2).0014)
- Koraishi, O., 2023. Teaching English in the Age of AI: Embracing ChatGPT to Optimize EFL Materials and Assessment. *Language Education and Technology*. [online] Available at <https://langedutech.com/letjournal/index.php/let/article/view/48> [Accessed 17 August 2023].
- Kostka, I., and Toncelli, R., 2023. Exploring Applications of ChatGPT to English Language teaching: Opportunities, challenges, and recommendations. *The Electronic Journal for English as a Second Language*. [online] Available at <<https://tesl-ej.org/wordpress/issues/volume27/ej107/ej107int/>> [Accessed 18 October 2023].
- Kumar, A., 2023. Analysis of ChatGPT Tool to Assess the Potential of its Utility for Academic Writing in Biomedical Domain. *BEMS Reports*, 9 (1), pp 24–30. <https://doi.org/10.5530/bems.9.1.5>
- Lecler, A., Duron, L., and Soyer, P., 2023. Revolutionizing radiology with GPT-based models: Current applications, future possibilities and limitations of ChatGPT. *Diagnostic and Interventional Imaging*, 104(6), pp 269–274. <https://doi.org/10.1016/j.diii.2023.02.003>
- Liu, L., 2023. Analyzing the text contents produced by ChatGPT: Prompts, feature-components in responses, and a predictive model. *Journal of Educational Technology Development and Exchange*, 16 (1), pp 49–70. <https://doi.org/10.18785/jetde.1601.03>
- Liu, L., and Gibson, D., 2023. Exploring the use of ChatGPT for learning and research: Content data analysis and concerns. Society for Information Technology & Teacher Education International Conference, New Orleans, Louisiana. [online] Available at <<https://www.learntechlib.org/p/221924/>> [Accessed 18 October 2023].
- Mallow, J., 2023. ChatGPT For Students: How AI Chatbots Are Revolutionizing Education. *eLearning Industry*. [online] Available at <<https://elearningindustry.com/chatgpt-for-students-how-ai-chatbots-are-revolutionizing-education>> [Accessed 12 September 2023].
- Ninaus, M. and Sailer, M., 2022. Closing the loop – The human role in artificial intelligence for education. *Frontiers in Psychology*, 13. <https://doi.org/10.3389/fpsyg.2022.956798>
- Opitz, B., Ferdinand, N.K., and Mecklinger, A., 2011. Timing matters: the impact of immediate and delayed feedback on artificial language learning. *Front Hum Neurosci*. doi:10.3389/fnhum.2011.00008
- Paris, B., 2022. Instructors' perspectives of challenges and barriers to providing effective feedback. *Teaching & Learning Inquiry: The ISSOTL Journal*, 10. <https://doi.org/10.20343/teachlearningqu.10.3>
- Pisan, Y., Sloane, A. M., Dale, R., and Richards, D., 2002. Providing timely feedback to large classes. *IEEE Xplore, International Conference on Computers in Education*. <https://doi.org/10.1109/CIE.2002.1185961>
- Randell, B., and Coghlan, B., 2023. ChatGPT's astonishing fabrications about Percy Ludgate. *IEEE Annals of the History of Computing*, 45 (2), pp 71–72. <https://doi.org/10.1109/mahc.2023.327298>
- Regional Convention on the Recognition of Studies, Diplomas and Degrees in Higher Education in Latin America and the Caribbean. 2019. UNESCO. [online] Available at < <https://www.unesco.org/en/legal-affairs/regional-convention-recognition-studies-diplomas-and-degrees-higher-education-latin-america-and-0> > [Accessed 11 July 2023].
- Tiili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., and 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), p 15. <https://doi.org/10.1186/s40561-023-00237-x>
- Vijayakumar, B., Höhn, S., and Schommer, C., 2019. Quizbot: Exploring Formative Feedback with Conversational Interfaces. *Communications in computer and information science*, pp 102–120. https://doi.org/10.1007/978-3-030-25264-9_8
- Wang, T. et al. (2023) 'Exploring the potential impact of artificial intelligence (AI) on international students in higher education: generative AI, chatbots, analytics, and international student success,' *Applied Sciences*, 13(11), p. 6716. <https://doi.org/10.3390/app13116716>