



Sheikhbahaee University

Foreign Language Teaching and Translation Studies



ORIGINAL RESEARCH ARTICLE

Comparative Analysis of AI vs. Human Feedback Effects on IELTS Candidates' Writing Performance

Alireza Ramazani; Houman Bijani *^{ID}; Mohammad Reza Oroji

Department of English Language, Islamic Azad University, Zanzan Branch, Zanzan, Iran

*Corresponding author: houman.bijani@gmail.com

doi 10.22034/efl.2025.493559.1334

ARTICLE INFO

Article History:

Received: 12 December 2024

Revised: 08 February 2025

Accepted: 27 April 2025

Keywords:

AI-Generated Feedback;
Assessment; Feedback
Comparison; Human
Feedback; IELTS
Writing; Writing
Performance

ABSTRACT

This study compares the effectiveness of artificial intelligence (AI)-generated feedback (specifically using ChatGPT) and human feedback in improving the writing performance of B2-level IELTS candidates. A mixed-methods approach was employed, combining quantitative and qualitative data to evaluate the impact of both types of feedback. A total of 100 B2-level English learners from various cities of Iran were selected through convenience sampling and then randomly assigned to two groups. Pre- and post-test assessments, using mock IELTS exams, provided quantitative data, while qualitative data were gathered through participant questionnaires. These findings reinforce the superiority of human feedback in enhancing higher-order writing skills; however, this study extends prior research by examining the effectiveness of AI-generated and human feedback within the specific context of Iranian IELTS candidates. Given the traditionally grammar-focused writing instruction in Iran, the results highlight the need for integrated feedback approaches that address coherence, cohesion, and argument development. The study also offers practical implications for IELTS preparation programs in non-native English-speaking countries, advocating for a structured combination of AI and human feedback to maximize writing proficiency. Learners valued the personalized guidance of human feedback, which fostered motivation and engagement. Conversely, AI-generated feedback excelled in correcting surface-level errors, such as grammar and vocabulary, offering immediacy and consistency. Nonetheless, AI's lack of interpretive depth limited its ability to address complex writing issues like argumentative development and cohesive flow. This study underscores the potential benefits of integrating AI and human feedback to enhance writing instruction for standardized assessments. Future research should investigate strategies to refine AI tools for better support of advanced writing skills.

How to cite this article: Ramazani, A., Bijani, H., & Oroji, M. R. (2025). Comparative analysis of AI vs. human feedback effects on IELTS candidates' writing performance. *Journal of Foreign Language Teaching and Translation Studies*, 10(1), 17–40. doi: 10.22034/efl.2025.493559.1334

2645-3592/ © 2025 The Authors. Published by Sheikhbahaee University Press.

This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>).



1. Introduction

In recent years, the integration of technology into education has revolutionized teaching and learning. This shift has been driven by the need for more flexible, accessible, and scalable educational solutions that cater to diverse learner needs. Artificial intelligence (AI) has become an integral part of language learning, particularly in providing automated feedback on writing tasks. AI-powered tools, such as ChatGPT, offer instant feedback on grammar, vocabulary, and coherence, making them a valuable resource for learners preparing for high-stakes exams like the International English Language Testing System (IELTS). However, despite the growing use of AI in language education, research on its effectiveness in IELTS writing preparation remains limited, particularly in non-Western contexts. Most existing studies focus on AI-generated feedback in general ESL/EFL writing, leaving a gap in understanding how AI-generated feedback compares to human feedback in the specific demands of IELTS writing. This study aims to address this gap by evaluating the impact of AI and human feedback on Iranian IELTS candidates' writing performance, offering insights into how these feedback types contribute to language proficiency in a high-stakes testing environment. The incorporation of digital tools and platforms has facilitated new pedagogical approaches, allowing for more dynamic, interactive, and personalized learning experiences. Among these technological advancements, artificial intelligence (AI) has emerged as a transformative tool, offering innovative methods for delivering feedback that were previously unattainable. As education evolves in the 21st century, the role of AI in providing feedback has garnered significant attention from both educators and researchers. This interest stems from a desire to enhance student outcomes, particularly in language learning, where feedback plays a pivotal role in skill development.

AI technologies, such as natural language processing (NLP) models and machine learning algorithms, have enabled the creation of systems capable of analyzing written texts and providing feedback almost instantaneously. These systems offer a range of functionalities, from correcting grammatical errors to suggesting improvements in structure and style. Such capabilities have opened up new possibilities for language learners, particularly those preparing for high-stakes exams like the International English Language Testing System (IELTS). In this context, understanding the comparative effectiveness of AI-generated feedback versus traditional human feedback has become a critical area of inquiry. While AI-generated feedback offers the potential for scalability and consistency, the nuanced and context-sensitive feedback provided by human educators is often seen as irreplaceable. This tension between AI and human feedback lies at the heart of ongoing research efforts aimed at improving language instruction and assessment.

Writing, as a foundational skill in education, holds particular importance in the realm of language acquisition. It is not merely a mechanical process of recording thoughts but a cognitively demanding activity that requires the integration of multiple linguistic and cognitive processes. As Grabe and Kaplan (1996) suggest, writing involves complex operations such as idea generation, organization, and revision, which must be synchronized with language-specific tasks like grammar, vocabulary selection, and the creation of coherent and cohesive text. In second language acquisition (SLA), writing is often one of the most challenging skills to master due to the need for learners to simultaneously focus on linguistic accuracy, stylistic appropriateness, and the effective communication of ideas. This complexity underscores the need for effective feedback, which serves as a crucial mechanism for improving learners' writing performance.

Feedback in language learning has long been recognized as a key driver of student improvement. It provides learners with insights into their current performance, highlighting both strengths and areas for improvement. Feedback is also a powerful instructional tool, guiding students through the iterative process of revising and refining their work. The definition by Hattie and Timperley (2007) outlined that feedback is "information provided by an agent (e.g., teacher, peer, book, parent, or experience) regarding aspects of one's performance or understanding" (p. 81). In the context of language learning, feedback can take many forms, including comments on grammatical accuracy, lexical choice, coherence, and the overall structure of the text. Research on the effectiveness of feedback has consistently shown that timely, specific, and constructive feedback can lead to significant improvements in student performance. Shute (2008) emphasizes that feedback is most effective when it is clear, targeted, and provides actionable suggestions that learners can implement in subsequent revisions of their work.

The IELTS exam, which stands as one of the most widely recognized and respected English proficiency tests globally, places a strong emphasis on writing proficiency. The IELTS Writing section is designed to evaluate a candidate's ability to produce clear, well-organized, and grammatically accurate text across two writing tasks. These tasks require candidates to demonstrate their ability to convey ideas effectively while adhering to specific genre conventions and responding appropriately to prompts. The IELTS Writing assessment is based on four key criteria: Task Achievement, Coherence and Cohesion, Lexical Resource, and Grammatical Range and Accuracy (IELTS, 2021). Each criterion captures a different aspect of writing performance, with Task Achievement focusing on the completeness and relevance of the response, Coherence and Cohesion evaluating the logical flow and connection of ideas, Lexical Resource assessing the range and appropriateness of vocabulary, and Grammatical Range and Accuracy measuring the diversity and correctness of sentence structures. Given the importance of writing in determining a candidate's overall band score, the provision of effective feedback is critical for improving performance in this section of the exam.

The advent of AI (e.g., Burgess & Lacy, 2018) has introduced new possibilities for providing feedback on writing, with tools like ChatGPT leading the way in offering automated responses to student work. Developed by OpenAI, ChatGPT is a sophisticated AI language model that generates human-like text based on the input it receives. It is capable of analyzing written texts and providing feedback on a range of issues, including grammar, coherence, lexical choice, and style. By simulating the responses of a human tutor, ChatGPT aims to assist learners in improving their writing through detailed feedback and suggestions for revision. One of the key advantages of AI tools like ChatGPT is their ability to provide immediate feedback, allowing learners to receive real-time insights into their work. This contrasts with traditional human feedback, which is often delayed due to the time constraints faced by instructors. Additionally, AI-generated feedback systems can be scaled to accommodate large numbers of learners, making them a potentially cost-effective solution for institutions seeking to enhance writing instruction.

Despite these advantages, the effectiveness of AI-generated feedback remains a topic of ongoing debate and research. While AI tools can offer quick and consistent feedback, they may lack the depth of understanding and contextual awareness that human teachers bring to the feedback process. Human feedback, according to the findings of the study by Wang and Brown (2020), is often valued for its ability to address not only surface-level errors but also deeper issues related to argumentation, clarity, and audience awareness. Moreover, human teachers can engage in meaningful dialogue with learners, offering personalized guidance that takes into account the learner's unique needs, strengths, and weaknesses. In contrast, AI-generated feedback, while efficient, may struggle to provide the same level of personalization and may overlook subtleties in the writing that a human reader would catch.

In the context of IELTS writing, the quality and type of feedback provided can have a significant impact on a student's performance. Given the high stakes associated with the exam, learners rely on feedback to identify weaknesses in their writing and make targeted improvements. While human feedback is traditionally regarded as the gold standard due to its personalized and context-aware nature, AI tools like ChatGPT offer compelling advantages in terms of accessibility, consistency, and affordability. As AI continues to develop, it is essential to understand how these two forms of feedback compare in their ability to enhance IELTS writing performance. Such an understanding is critical not only for educators and learners but also for policymakers and institutions seeking to integrate AI into language learning programs.

Given the context outlined, the research questions are defined as follows:

1. What is the comparative impact of AI-generated feedback and human feedback on IELTS writing performance, specifically in terms of Coherence and Cohesion, Lexical Resource, Grammatical Range and Accuracy, and Task Achievement?
2. How do students perceive the usefulness, accuracy, and reliability of AI-generated feedback compared to human feedback in improving IELTS writing performance?
3. What is the impact of AI-generated feedback versus human feedback on specific aspects of IELTS writing performance?"

2. Literature Review

The literature review will cover a historical overview of feedback in language education, the development of writing feedback strategies, and the emergence of technology-enhanced feedback systems, especially focusing on AI-based tools like ChatGPT including recent Iranian studies and their contributions to understanding feedback mechanisms in the context of language learning, particularly within high-stakes environments like the International English Language Testing System (IELTS).

2.1. Theoretical Background of Feedback in Language Learning

Feedback has been a central concept in language learning for decades. Initially, behaviorist theories (e.g., Skinner, 1954) proposed that feedback should be immediate and used as a reinforcement mechanism to condition desired responses. However, cognitive theories, such as those proposed by Piaget (1970), emphasized the importance of internal cognitive processes, suggesting that feedback should be more focused on guiding learners' self-discovery and promoting active engagement with the learning process.

Vygotsky's (1978) socio-cultural theory introduced the concept of the Zone of Proximal Development (ZPD), which framed feedback as a tool that helps learners perform tasks just beyond their current level of competence, with guidance from more knowledgeable others. This perspective influenced subsequent models of feedback, emphasizing its interactive nature and its role in fostering cognitive development.

In writing instruction, feedback has been shown to significantly influence the development of writing skills. It helps learners identify errors, improve linguistic accuracy, and refine their ability to organize ideas, structure arguments, and maintain coherence in their writing. While there is no one-size-fits-all approach to feedback, research suggests that feedback is most effective when it is timely, constructive, and addresses both surface-level and higher-order concerns in writing (Hyland & Hyland, 2006).

2.2. Characteristics of Effective Human Feedback

Human feedback in language learning is widely recognized for its effectiveness in addressing higher-order writing skills, such as coherence, argumentation, and task fulfillment. Unlike AI-generated feedback, which often focuses on surface-level corrections (e.g., grammar and vocabulary), human feedback offers in-depth explanations, personalized guidance, and interactive clarification, all of which contribute to deeper learning (Hyland & Hyland, 2006). This aligns with Vygotsky's (1978) socio-cultural theory, which emphasizes the importance of scaffolding in advancing learners' cognitive skills. Through direct engagement with an instructor, learners can better understand how to refine their ideas, improve cohesion, and develop a stronger argumentative structure.

However, the effectiveness of human feedback is also influenced by cultural and educational contexts. In Iran, teacher-centered instruction remains dominant, with students often viewing instructors as the primary source of authoritative knowledge (Zarei & Sadighi, 2011). This dynamic may make Iranian IELTS candidates more receptive to human feedback than AI-generated feedback, which lacks direct human interaction. Moreover, prior research suggests that Iranian EFL learners tend to value detailed, teacher-provided feedback as a sign of instructional care and investment in their success (Vahdat & Khodabakhsh, 2019). Consequently, while AI-generated feedback provides immediacy and consistency, its lack of interpersonal engagement may limit its effectiveness in contexts where teacher feedback is deeply ingrained in the learning process. Another key characteristic of effective human feedback is its timeliness. Shute (2008) argued that feedback is most effective when it is provided shortly after the learner's performance, allowing the learner to immediately apply the feedback to improve their language skills. This is particularly important in language learning, where timely feedback can help learners correct errors before they become ingrained.

Hyland and Hyland (2006) also highlighted the importance of providing feedback that is constructive and supportive. Feedback that is overly critical or negative can demotivate learners, while feedback that is framed positively can encourage learners to persevere and continue improving their language skills.

2.3. Challenges and Limitations of Human Feedback

Despite its many benefits, such as providing immediate feedback, consistency and objectivity, developing learner autonomy, accessibility just to name a few, human feedback in language learning also has several limitations. One of the primary challenges is the subjectivity inherent in human judgment. Research by Bitchener, Young, and Cameron (2005) found that different instructors often provide different feedback on the same piece of writing, leading to inconsistencies in the feedback that learners receive. This can be particularly problematic in standardized testing contexts, where consistent feedback is essential for ensuring fairness.

Another challenge is the time and effort required to provide personalized feedback to each learner. In large language classes, it can be difficult for instructors to provide the detailed, individualized feedback that is necessary for effective language learning. Ferris (2003) highlighted the challenges of providing written feedback in ESL (English as a Second Language) classrooms, where instructors often struggle to provide timely and detailed feedback to all students.

2.4. Writing in the Iranian Context

Iranian EFL instruction has long been characterized by a strong emphasis on teacher-centered methodologies, where explicit grammar instruction, rote memorization, and instructor authority play dominant roles (Zarei & Sadighi, 2011). These educational traditions shape students' expectations for feedback, often leading them to favor detailed, directive guidance from instructors over self-guided revisions. Such an approach has implications for the effectiveness of AI-generated versus human feedback in IELTS writing preparation.

Human feedback aligns closely with Iranian students' expectations, as it offers direct, authoritative explanations and structured guidance, reinforcing their traditional learning experiences. Iranian learners may perceive teacher-provided feedback as more credible, reliable, and actionable, particularly when addressing higher-order writing concerns such as argument development and coherence (Vahdat & Khodabakhsh, 2019). In contrast, AI-generated feedback, which lacks interpersonal interaction and pedagogical scaffolding, may be viewed as impersonal and insufficient for tackling complex writing tasks. Additionally, learners accustomed to teacher-directed learning may struggle to interpret and apply AI-generated feedback independently, further limiting its effectiveness in this context.

Despite these challenges, AI-generated feedback provides unique benefits, such as immediacy and consistency, which may be advantageous in Iranian classrooms with large student populations where personalized teacher feedback is limited. However, for AI-generated feedback to be effective in this setting, it may need to be supplemented with teacher explanations or integrated into blended learning models that balance automation with human interaction. Understanding these cultural dynamics is crucial for developing effective feedback mechanisms that align with Iranian EFL learners' educational backgrounds and expectations.

Iranian learners face specific challenges in academic writing, including a limited understanding of genre conventions and an over-reliance on memorization rather than critical analysis (Moinzadeh & Zafarghandi, 2015). Research by Rahimi and Kafipour (2014) shows that many Iranian students struggle with constructing coherent arguments, which affects their performance in high-stakes assessments like IELTS. This highlights the necessity for targeted writing instruction that addresses both linguistic skills and the cognitive strategies required for effective writing.

Iranian studies emphasize the importance of context in writing instruction. For instance, Zarei and Sadighi (2011) conducted a study on IELTS writing challenges, revealing that Iranian EFL learners often lack familiarity with the test's expectations, resulting in lower performance. Additionally, Maleki and Zare (2015) found that tailored preparatory courses focusing on IELTS writing tasks significantly improve students' confidence and scores.

Furthermore, Vahdat and Khodabakhsh (2019) explored the impact of feedback on Iranian learners' writing, highlighting the effectiveness of both peer and teacher feedback in enhancing writing quality. These studies

underscore the need for comprehensive writing instruction that incorporates culturally relevant pedagogical strategies to address the specific needs of Iranian learners.

2.5. The Role of Technology in Feedback Provision

Advancements in educational technology have significantly impacted feedback provision in language learning. The introduction of Computer-Assisted Language Learning (CALL) in the 1960s revolutionized language teaching, allowing for more personalized and immediate feedback. The development of Intelligent Tutoring Systems (ITS) and Natural Language Processing (NLP) tools has further advanced this trend, enabling real-time feedback on both form and content (Burgess & Lacy, 2018).

The integration of AI in language education, particularly in writing tasks, has garnered attention due to its ability to analyze text at a granular level. Tools like Grammarly, Criterion, and ChatGPT have become widely used for providing automated feedback on grammar, style, and structure. These tools employ AI and NLP technologies to assess written work and offer corrective suggestions instantly, thus removing the delay typically associated with human feedback. While AI-driven feedback systems offer efficiency and scalability, their implementation in educational settings raises several ethical concerns. One major issue is data privacy, as AI tools often require users to upload writing samples, potentially exposing personal information. Without robust data protection measures, learner data may be stored, shared, or even misused by third-party platforms, raising concerns about confidentiality and security (Elliot et al., 2021).

Additionally, algorithmic bias presents another challenge. Many AI-generated feedback tools are trained on large-scale datasets that predominantly reflect Western academic writing conventions. As a result, non-Western learners, including Iranian IELTS candidates, may receive feedback that does not fully align with their linguistic background or rhetorical traditions. This can lead to unintentional disadvantages, where AI disproportionately favors certain sentence structures or lexical choices that conform to dominant English norms (Wang & Brown, 2020).

Furthermore, over-reliance on AI-generated feedback may discourage learners from developing independent revision skills. While AI provides immediate corrections, students may become passive recipients of automated suggestions rather than actively engaging in the writing process. This concern highlights the need for a balanced approach, where AI-generated feedback is integrated into a broader pedagogical framework that fosters critical thinking and learner autonomy (Ranalli et al., 2022). Studies by Almarwani et al. (2021) and Shirvani and Shirmohammadi (2020) have demonstrated that AI-driven feedback systems can significantly enhance the writing performance of learners, particularly in large classroom settings where individual attention from human instructors is limited.

AI-based feedback systems, however, are not without limitations. While they are adept at identifying surface-level errors (such as grammar and punctuation), they are less capable of providing feedback on higher-order concerns such as coherence, cohesion, and argumentation. This limitation underscores the continued need for human feedback, particularly in tasks that require nuanced understanding and critical engagement with ideas (Hyland & Hyland, 2019). In Iranian contexts, AI tools have been found effective for immediate corrections but are often supplemented by teacher feedback for more comprehensive learning outcomes (Shirvani & Shirmohammadi, 2020). A study by Wang and Brown (2020) compared the effectiveness of AI-generated feedback with human feedback in an ESL writing course. Their findings indicated that students who received AI-generated feedback showed greater improvements in areas such as grammar and coherence than those who relied solely on traditional teacher feedback.

2.6. Effectiveness of AI-generated feedback

Several studies have demonstrated that AI-generated feedback can be as effective as, or more effective than human feedback in certain areas of language learning. One of the key advantages of AI-generated feedback is its consistency and objectivity. Unlike human feedback, which can be subjective and vary between instructors, AI-generated feedback is standardized and based on pre-programmed rules, ensuring that all learners receive the same

level of feedback. This consistency is particularly important in standardized testing contexts, such as the IELTS Writing section, where uniformity in feedback can contribute to fairness and reliability.

A study by Almarwani, Jones, and Dray (2021) found that students who used AI-generated feedback tools, such as Grammarly, improved their writing performance in terms of grammatical accuracy and lexical variety. However, the study also highlighted the limitations of AI-generated feedback, particularly in addressing higher-order concerns such as coherence, argumentation, and creativity. While AI tools excel at correcting surface-level errors, they may struggle to provide meaningful feedback on the overall structure and content of the text. Also, a study by Wang and Brown (2020) compared the impact of AI-generated feedback with traditional teacher feedback in an ESL writing course. The study found that students who received AI-generated feedback showed greater improvements in their writing, particularly in areas such as grammar and coherence.

In another study, Zhang and Hyland (2020) examined learner perceptions of AI-generated feedback in an EFL writing course. Their research found that while many learners appreciated the immediacy and objectivity of AI-generated feedback, they also expressed concerns about the lack of personalization and the inability of AI systems to provide in-depth feedback on more complex aspects of writing. These findings suggest that while AI-generated feedback can be a valuable tool in language learning, it is most effective when used in conjunction with human feedback.

2.7. Learner Perceptions of AI-generated feedback

Learner perceptions of AI-generated feedback are crucial for its effectiveness. Studies have shown that while some learners appreciate the immediacy and objectivity of AI-generated feedback, others are skeptical of its ability to understand the nuances of their writing. A study by Zhang and Hyland (2020) found that while students generally found AI-generated feedback helpful, they also expressed concerns about its lack of personalization and the absence of human interaction.

Learner perceptions of AI-generated feedback vary by language proficiency level. For example, beginners may find AI-generated feedback more helpful because it focuses on basic errors, while advanced learners may find it less useful for developing more sophisticated language skills. In their study on the use of AI in language learning, Huang, Chen, and Lin (2021) found that advanced learners were more critical of AI-generated feedback, particularly when it came to its ability to provide nuanced and context-specific suggestions.

A study conducted in Iran by Farhadi and Jafari (2019) similarly found that learners who received teacher feedback reported higher levels of motivation and engagement compared to those who relied solely on AI tools for feedback. The learners expressed that human feedback was more meaningful, as it provided them with detailed explanations and emotional encouragement, whereas AI-generated feedback was perceived as mechanical and overly focused on error correction. The researchers concluded that the emotional component of feedback plays a critical role in shaping student attitudes toward their writing development, especially in EFL contexts where learners may struggle with confidence and self-efficacy.

2.8. Comparative Studies: AI vs. Human Feedback

Comparative studies on AI-generated feedback and human feedback have provided valuable insights into the strengths and weaknesses of each approach. One of the earliest comparative studies was conducted by Ware and Warschauer (2006), who analyzed the impact of digital writing feedback on student performance. Their research showed that while digital feedback tools, including AI-driven systems, were effective in improving technical aspects of writing, human feedback was more effective in fostering critical thinking and creativity.

Lu and Law (2012) conducted a meta-analysis of studies comparing computer-assisted feedback with traditional feedback methods in language learning. Their analysis revealed that while AI-generated feedback was generally more effective in improving surface-level language skills, human feedback was superior in addressing more complex aspects of language use, such as discourse and pragmatics.

More recent studies have focused on the integration of AI-generated feedback into hybrid learning environments, where students receive both AI and human feedback. For instance, Xie, Chu, Hwang, and Wang (2019) explored the use of AI tools in a blended learning environment and found that students who received both types of feedback performed better overall than those who received only one type. This suggests that a combination of AI and human feedback may be the most effective approach to language learning.

An increasing number of studies have compared the effectiveness of AI-generated and human feedback across educational contexts, particularly in writing instruction. Research by Warschauer and Grimes (2008) demonstrated that AI-generated feedback systems like Criterion provided students with immediate corrective feedback on surface-level errors such as grammar and spelling. However, their study found that AI tools were less effective in addressing higher-order concerns such as idea development, organization, and critical thinking. In contrast, human feedback was more successful in promoting cognitive engagement and providing the nuanced, context-specific guidance that students require for more complex writing tasks. Warschauer and Grimes (2008) concluded that while AI-generated feedback systems offer significant advantages in terms of speed and consistency, they should be used as a supplement rather than a replacement for human feedback, especially in high-stakes writing contexts like the IELTS exam.

2.9. The Role of Feedback in IELTS Preparation

Effective feedback is crucial for improving performance in the IELTS Writing section. Research has shown that targeted feedback, which addresses specific areas for improvement, can significantly enhance a candidate's writing skills. For example, Green's (2007) study on IELTS preparation found that students who received detailed feedback on their writing were more likely to achieve higher scores on the Writing section. His research emphasized the importance of providing feedback that addresses specific IELTS writing criteria, such as grammatical accuracy and lexical resource.

Taylor and Weir (2012) examined the role of feedback in IELTS writing performance, finding that students who received ongoing formative feedback throughout their preparation demonstrated significant improvements in their overall writing scores. The study highlighted the importance of regular feedback that targets specific weaknesses, such as grammatical errors, task achievement, and coherence and cohesion, in helping students meet the IELTS Writing assessment criteria.

Research on feedback in IELTS preparation highlights the importance of focused corrective feedback, particularly in addressing recurrent grammatical errors and improving task achievement (Hyland & Hyland, 2006). Studies suggest that a combination of teacher feedback and peer review can be effective in helping learners refine their writing, especially when the feedback is aligned with the IELTS Writing band descriptors (Ferris, 2009).

AI tools like ChatGPT, which provide immediate feedback on writing, are increasingly being integrated into IELTS preparation programs. However, the effectiveness of AI-generated feedback in this context remains a topic of debate. Studies like that of Almarwani et al. (2021) have demonstrated that AI-generated feedback tools can be beneficial in helping students improve their grammatical accuracy and lexical range. However, other researchers argue that the limitations of AI-generated feedback—particularly its inability to provide in-depth comments on argument development and coherence—make it insufficient for preparing students for the more complex writing tasks required in IELTS.

Additionally, AI tools like Grammarly and Criterion have been found to be particularly effective in providing immediate feedback on grammar and style, which are essential components of the IELTS Writing section. Studies by O'Sullivan and Weir (2011) demonstrated that learners who used AI tools in conjunction with human feedback showed greater improvements in their IELTS Writing scores compared to those who relied solely on traditional feedback.

In contrast, human feedback remains critical for improving more complex aspects of writing, such as coherence and cohesion. O'Sullivan and Green (2003) argue that while AI tools can help learners improve technical accuracy,

human instructors are better equipped to provide feedback on how to organize ideas, structure arguments, and develop a clear line of reasoning—all key elements of the IELTS Writing section.

In the Iranian context, research by Zarei and Hashemnezhad (2020) focused on the use of AI tools in IELTS preparation among Iranian EFL learners. The study found that students who received human feedback, either from teachers or peers, performed better in writing tasks that required higher-order cognitive skills, such as developing and organizing ideas, compared to those who relied on AI-generated feedback. The researchers concluded that while AI tools offer valuable support in terms of grammar and lexical feedback, they should be supplemented with human feedback to ensure students are fully prepared for the complex writing tasks required in IELTS.

3. Methodology

3.1. Research Design

This study employs a mixed-methods approach. This approach combines quantitative and qualitative data collection methods to achieve a comprehensive understanding of the research question. This study employed a mixed-methods approach, integrating quantitative (pre-test and post-test scores) and qualitative (participant questionnaires) data to provide a comprehensive evaluation of feedback effectiveness. The quantitative data measure objective performance improvements, while the qualitative insights capture learner perceptions, engagement, and preferences, enriching the understanding of feedback impact. Regarding feedback structure, AI-generated feedback was standardized, providing consistent grammatical, lexical, and coherence-related corrections across all writing tasks. In contrast, human feedback was personalized, with instructors offering task-specific guidance, tailored explanations, and interactive clarification based on individual learner needs. A pre-test and post-test design was implemented to quantitatively assess the participants' performance across the four IELTS writing criteria, allowing for a direct comparison of results before and after the feedback intervention. The pre-test established a baseline measure of writing proficiency, while the post-test evaluated improvements resulting from the feedback provided.

In addition to the quantitative assessment, a qualitative questionnaire was administered to gather insights into participants' perceptions of the feedback they received. This qualitative component is crucial for understanding the subjective experiences of learners and their preferences regarding AI-generated versus human feedback. By employing this mixed-methods design, the study seeks to triangulate data and enhance the validity of the findings.

3.2. Participants

The participants in this study were selected through convenience sampling, aimed at obtaining a representative sample of B2-level English learners across various cities in Iran. This study employed convenience sampling, selecting participants from readily accessible IELTS candidates across various cities in Iran. While this approach facilitated efficient data collection, it introduces potential selection bias, as participants may not be fully representative of the broader population of IELTS learners. As a result, the findings may be less generalizable to diverse learner demographics, such as candidates with different proficiency levels or those in different educational settings. Future studies should consider random sampling techniques to enhance the representativeness of the sample and strengthen the external validity of the results. Another key limitation is the absence of a control group, which makes it difficult to fully isolate the effects of AI-generated and human feedback from other influencing variables, such as prior writing experience, individual motivation, or external instruction. While the pre-test and post-test design allows for a comparative analysis of feedback impact, a control group receiving no intervention or a different form of feedback (e.g., peer feedback) would provide a clearer baseline for measuring effectiveness. Future research should incorporate randomized controlled trials (RCTs) or matched control groups to more accurately assess the causal effects of different feedback types on writing performance. Despite these limitations, the study provides valuable insights into the comparative impact of AI and human feedback on IELTS writing. By acknowledging these methodological constraints, we encourage further research that refines sampling strategies and experimental design to build upon the findings presented here. A total of 100 participants were recruited who were selected according to Cochran's formula of sample-size selection. Of the participants, 55% identified as female and 45% as male. The age range of participants varied significantly, encompassing young adults to older learners (13 to

22 years of age), thus representing a diverse range of backgrounds and learning experiences. The participants were all Iranian EFL learners doing their education at either high school or studying for their bachelor's at university.

All participants were classified as B2-level learners based on a standardized placement test administered prior to the study. This classification ensured that participants possessed the requisite language proficiency for engaging with IELTS writing tasks. Furthermore, participants were informed of the voluntary nature of the study and the option to withdraw at any time without penalty. Each participant was then randomly assigned to one of two groups: the AI-generated feedback group, which received feedback from ChatGPT, and the human feedback group, which received feedback from IELTS-trained instructors.

3.3. Instruments

This study utilized multiple instruments to collect data, ensuring a comprehensive evaluation of writing performance and participant perceptions:

1. ChatGPT (version 3.0): The AI-generated feedback was delivered using OpenAI's ChatGPT, which provided detailed feedback on participants' IELTS writing tasks. This feedback focused on the four key criteria of the IELTS writing assessment: Task Achievement, Coherence and Cohesion, Lexical Resource, and Grammatical Range and Accuracy. The AI system analyzed writing samples and offered suggestions for improvement based on established writing standards.

2. IELTS Writing Rubric: The standardized IELTS Writing Rubric was employed to evaluate participants' pre-test and post-test scores. This rubric provided a consistent framework for assessing writing performance across all participants.

3. Tehran IELTS Mock Exam: A standardized mock IELTS exam was administered as both the pre-test and post-test assessment. This exam served as a reliable tool for measuring participants' writing performance in a simulated testing environment.

4. Placement Test: A standardized placement test was administered to confirm the participants' B2-level proficiency in English before the study commenced. This test assessed participants' overall language skills, including reading, writing, listening, and speaking. The placement test results ensured that all participants had a foundational level of proficiency necessary for engaging with IELTS writing tasks. It is noteworthy to indicate that the reliability of the questionnaire was measured 0.78 using alpha Cronbach which is high and the validity using Confirmatory Factor Analysis (CFA) was measured to be within the acceptable range.

5. Questionnaire: A qualitative questionnaire was developed to gather participants' perceptions of the feedback received. This instrument included both open-ended and closed-ended questions designed to capture learners' preferences, insights into the feedback process, and reflections on their writing development. The qualitative questionnaire was designed to assess participants' perceptions of AI-generated and human feedback in terms of usefulness, accuracy, and engagement. It included a mix of open-ended and Likert-scale questions to capture both subjective insights and measurable trends. To ensure clarity and relevance, the questionnaire underwent a pilot test with a small group of 10 IELTS candidates, allowing for refinements based on participant feedback. Expert validation was conducted by two applied linguistics specialists, who reviewed the questions for content validity and alignment with the study's objectives.

3.4. Procedure

The study was conducted over six weeks and aimed to evaluate the impact of AI-generated and human feedback on participants' IELTS writing performance. In the first week, participants completed a standardized placement test to confirm their English proficiency at the B2 level, ensuring a uniform starting point for all participants. During the second week, all participants undertook the Tehran IELTS Mock Writing Exam, a standardized test designed to simulate real IELTS writing conditions. The pre-test was scored using the IELTS Writing Rubric, which assessed

performance across the four IELTS writing criteria: Task Achievement, Coherence and Cohesion, Lexical Resource, and Grammatical Range and Accuracy. These scores served as the baseline for comparison. After the pre-test, participants were randomly assigned to one of two groups: the AI-generated feedback group or the human feedback group.

From the third to the fifth week, both groups participated in a series of 15 feedback sessions, averaging five sessions per week. Each group completed a series of IELTS writing tasks, with writing submissions being assessed and feedback provided after every task. The AI-generated feedback group received detailed feedback generated by ChatGPT, designed to align with IELTS criteria and offer actionable suggestions for improvement. Similarly, the human feedback group received feedback from an IELTS-trained instructor, which was also tailored to meet the specific requirements of the IELTS Writing Rubric. These interactive feedback sessions allowed participants to ask questions and clarify feedback points, promoting a better understanding of their strengths and areas for improvement. Feedback was provided in individual sessions, allowing for personalized guidance and clarification. For AI-generated feedback, participants reviewed system-generated comments and had the option to discuss unclear points with the researcher. Human feedback sessions included written annotations followed by brief consultations to ensure understanding and application. To facilitate active engagement, students were required to revise their drafts based on feedback and submit improvements within 48 hours. During week six, a post-test assessed writing improvements, followed by a brief reflection survey to gather participant insights on the feedback process.

In the sixth week, all participants completed a second Tehran IELTS Mock Writing Exam as the post-test, which was scored using the same rubric to ensure consistency. The pre-test and post-test scores were compared to measure the effectiveness of the feedback received in improving participants' writing performance. In addition, participants completed a qualitative questionnaire at this stage to provide insights into their experiences with both feedback types. The questionnaire included a combination of open-ended and closed-ended questions to gather comprehensive data about their perceptions and preferences.

Quantitative data were derived from the pre-test and post-test scores, while qualitative data were collected from the questionnaire, offering deeper insights into participants' experiences.

3.5. Data Analysis

The quantitative data collected from the pre-test and post-test exams were analyzed using SPSS (Version 22). Descriptive statistics, including mean scores and standard deviations, were calculated for each group to provide an overview of performance. An independent samples t-test was conducted to compare the mean scores of the AI-generated feedback group and the human feedback group, determining whether there were statistically significant differences in writing performance between the two groups.

For the qualitative data, a thematic analysis was employed. Key themes were identified from the open-ended questionnaire responses to explore participants' perceptions of the feedback they received and its impact on their writing performance. This analysis provided insights into participants' preferences, satisfaction levels, and perceived effectiveness of the feedback.

By combining quantitative statistical analysis with qualitative thematic analysis, the study adopted a mixed-methods approach to gain a comprehensive understanding of the effectiveness of AI-generated and human feedback. This dual approach enriched the findings, offering both numerical evidence and personal insights into how different types of feedback influenced IELTS writing performance.

To ensure the reliability and validity of the findings, measures such as scoring consistency, inter-rater reliability, and statistical analysis were implemented. Alpha Cronbach was used to measure the reliability of the questionnaire data and CFA was used to ascertain its construct validity. Trained evaluators scored the writing tasks, and inter-rater reliability was established by comparing scores from multiple raters on a subset of tasks. Both AI-generated and human feedback were analyzed for specificity and adherence to IELTS criteria, ensuring that the guidance

provided to participants was clear and actionable. Statistical methods were applied to the pre-test and post-test scores to account for confounding variables, such as prior writing experience or individual learning styles.

4. Results

Quantitative Findings

3.6. Pre-test and Post-test Results

The quantitative analysis evaluated participants' writing performance based on pre-test and post-test scores for both feedback groups. The results indicate that human feedback led to greater lexical improvement compared to AI-generated feedback. This finding aligns with previous research (Hyland & Hyland, 2006; Zhang & Hyland, 2020) but requires further exploration. One key reason for the superiority of human feedback in lexical development is its ability to provide nuanced word choice recommendations. Human instructors offer context-sensitive suggestions that extend beyond mere correctness, helping students refine their vocabulary based on tone, meaning, and appropriateness within an academic or argumentative writing context. Additionally, human feedback is particularly effective in addressing collocational accuracy—an essential component of lexical resource in IELTS writing. While AI-generated feedback can highlight isolated vocabulary errors, it often fails to recognize unnatural word pairings or suggest idiomatic and academic collocations that enhance fluency. In contrast, human instructors provide targeted feedback on natural word combinations and offer alternative phrasing to ensure lexical appropriateness. Furthermore, instructors assist learners in developing stylistic awareness, adjusting vocabulary based on formality, genre, and rhetorical purpose. For instance, IELTS writing requires a balance between formal academic language and clear argumentation, a distinction that human feedback effectively addresses through explicit explanations and examples. AI-generated feedback, while efficient in identifying repetitive or basic errors, lacks the interpretive depth needed to guide students in making strategic lexical choices for different writing contexts. Finally, the interactive nature of human feedback fosters learner engagement and vocabulary retention. Students can ask instructors for clarifications, synonyms, and alternative expressions, allowing for a more dynamic learning process that AI-generated feedback cannot replicate. These advantages explain why human feedback plays a critical role in enhancing lexical diversity, precision, and overall writing quality among IELTS candidates. The results are given in Table 1.

Table 1

Comparison of Pre-test and Post-test Mean Scores between AI and Human Feedback Groups

| Group | Pre-test Mean Score (\pm SD) | Post-test Mean Score (\pm SD) | Improvement (Mean Difference) |
|-----------------------------|------------------------------------|-------------------------------------|----------------------------------|
| AI-generated feedback Group | 5.5 \pm 0.74 | 6.5 \pm 0.74 | 1.0 |
| Human Feedback Group | 5.5 \pm 0.71 | 6.8 \pm 0.71 | 1.3 |

The human feedback group showed a greater improvement in post-test scores (6.8) compared to the AI feedback group (6.5). Both groups maintained their respective levels of variability in post-test scores, as indicated by unchanged SD values. The AI-generated feedback group showed an average improvement of 1.0 band score, while the human feedback group exhibited a higher improvement of 1.3 band score. The greater improvement in the human feedback group underscores its superior effectiveness in enhancing writing performance, as further reflected in the criterion-level analysis.

Table 2

Independent Samples T-Test Results

| Group Comparison | t-value | df | p-value |
|------------------|---------|----|---------|
|------------------|---------|----|---------|

| | | | |
|---|------|----|-------|
| AI-generated feedback vs. Human Feedback | 2.07 | 98 | 0.041 |
|---|------|----|-------|

A t-value of 2.07 was yielded by the independent samples t-test with a *df* of 98 and a p-value of 0.041, indicating a statistically significant difference between the two groups ($p < 0.05$). These results confirm that human feedback has a greater overall impact on improving writing performance compared to AI-generated feedback.

To provide a nuanced understanding, we analyzed the improvements across the four key IELTS writing criteria: Task Achievement, Coherence and Cohesion, Lexical Resource, and Grammatical Range and Accuracy. Each criterion was assessed to evaluate the effectiveness of each feedback type.

Table 3

Score Improvements per Criterion

| Writing Criteria | AI-generated feedback Mean Increase | Human Feedback Mean Increase | Average Score (Post-test) AI | Average Score (Post-test) Human |
|---------------------------------------|--|-------------------------------------|-------------------------------------|--|
| Task Achievement | 0.7 | 1.1 | 6.2 | 6.6 |
| Coherence and Cohesion | 0.6 | 1.05 | 6.1 | 6.5 |
| Lexical Resource | 0.8 | 1.3 | 6.3 | 6.7 |
| Grammatical Range and Accuracy | 0.9 | 1.4 | 6.4 | 6.8 |

The results from Table 3 provide a clear comparison of the strengths and limitations of AI and human feedback in improving IELTS Writing criteria. Human feedback demonstrated a more significant impact across all criteria, particularly in higher-order skills, while AI-generated feedback excelled in surface-level corrections. These findings reinforce the complementary roles of both feedback types and advocate for hybrid feedback systems to maximize learning outcomes.

Key Observations and Analysis

1. Task Achievement (TA)

Results: The greater increase in Task Achievement for the human feedback group (1.1) suggests a more effective approach to meeting task-specific requirements compared to the AI feedback group (0.7).

Interpretation: This reflects the ability of human feedback to address task-specific requirements comprehensively. Human instructors provided detailed guidance on understanding prompts, structuring responses, and meeting task expectations.

2. Coherence and Cohesion (CC)

Results: The human feedback group showed a higher mean increase (1.05) compared to the AI-generated feedback group (0.6).

Interpretation: Human feedback was more effective in improving the logical flow and connection of ideas. Learners benefited from tailored suggestions on paragraphing, transitions, and argument structuring, whereas AI-generated feedback provided more generic cohesion improvements.

3. Lexical Resource (LR)

Results: The human feedback group achieved a greater mean increase (1.3) than the AI-generated feedback group (0.8).

Interpretation: The significant improvement in the human feedback group highlights the nuanced vocabulary guidance provided by instructors, including suggestions for collocations, stylistic adjustments, and contextually appropriate word choices. AI-generated feedback, while helpful for identifying simple errors, was less effective in promoting lexical variety.

4. Grammatical Range and Accuracy (GRA)

Results: Both groups demonstrated notable improvements, with the human feedback group showing a greater increase (1.4) compared to the AI-generated feedback group (0.9).

Interpretation: AI-generated feedback effectively addressed surface-level grammatical errors, such as subject-verb agreement and tense consistency, but lacked the depth required for complex structures. Human feedback offered detailed explanations and corrections, enabling learners to master advanced grammar.

Overall Trends and Implications

Human Feedback's Superiority:

The human feedback group consistently outperformed the AI-generated feedback group across all criteria. This underscores the importance of personalized, detailed feedback in addressing higher-order skills and providing comprehensive support.

AI-generated feedback's Role:

AI-generated feedback showed meaningful contributions, particularly in GRA and LR, reflecting its strength in addressing technical aspects of writing. However, its limitations in addressing nuanced skills like coherence and argument development were evident.

Table 4

T-Values and P-Values for Each Writing Criterion

| Writing Criteria | AI-generated feedback Mean Increase | Human Feedback Mean Increase | T-Value | P-Value | Significance |
|--------------------------------|-------------------------------------|------------------------------|---------|---------|--------------|
| Task Achievement | 0.7 | 1.1 | 2.04 | 0.044 | yes |
| Coherence and Cohesion | 0.6 | 1.05 | 3.10 | 0.005 | yes |
| Lexical Resource | 0.8 | 1.3 | 2.37 | 0.020 | yes |
| Grammatical Range and Accuracy | 0.9 | 1.4 | 2.58 | 0.011 | yes |

The t-values and p-values across the four IELTS Writing criteria—Task Achievement (TA), Coherence and Cohesion (CC), Lexical Resource (LR), and Grammatical Range and Accuracy (GRA)—demonstrate statistically significant differences between the AI-generated feedback and human feedback groups. These results highlight the strengths and limitations of both feedback types while reinforcing the comparative effectiveness of human feedback.

Human feedback consistently led to greater improvements across all criteria, with its most pronounced impact on Coherence and Cohesion ($t = 3.10$, $p = 0.005$) and Grammatical Range and Accuracy ($t = 2.58$, $p = 0.011$). This suggests that the personalized nature of human feedback provided more tailored guidance, addressing higher-order skills like organizing ideas, using cohesive devices, and improving complex grammatical structures. Similarly, in Task Achievement ($t = 2.04$, $p = 0.044$) and Lexical Resource ($t = 2.37$, $p = 0.020$), human feedback helped participants refine their responses to prompts, enhance their vocabulary, and develop more varied lexical expressions. AI-generated feedback, on the other hand, showed significant contributions to surface-level corrections, particularly in grammar and vocabulary. However, its limitations were evident in addressing complex issues like

logical structuring and cohesion. While AI-generated feedback improved Grammatical Range and Accuracy (0.9 mean increase) and Lexical Resource (0.8 mean increase) effectively, its lack of depth in providing interpretive or task-specific guidance explains its relatively lower impact on higher-order skills compared to human feedback. Overall, the results support the complementary strengths of both feedback types. Human feedback excels in complex, context-specific areas of writing, while AI-generated feedback provides immediate, consistent support for foundational language skills. Together, they underline the potential of a blended feedback approach, combining AI's scalability with human instructors' depth, to optimize writing development for high-stakes assessments like the IELTS.

4.2. Qualitative Findings

Qualitative data were collected through a structured questionnaire designed to gauge participants' perceptions of the utility, effectiveness, and overall satisfaction with the feedback received. This section explores participants' insights, particularly regarding immediacy, efficiency, and the overall relevance of the feedback in their learning processes.

4.2.1. Perceptions of AI vs. Human Feedback

General Insights:

≠ AI-generated feedback Perception:

- 70% of participants reported finding AI-generated feedback helpful for identifying grammatical errors. Comments reflected on its strengths:
 - "I liked that the AI gave immediate feedback on my spelling mistakes, which helped me correct them on the spot."
 - "The rapid turnaround of AI-generated feedback was great, but sometimes I felt like it missed the nuances."
- However, 55% articulated concerns about AI's efficiency in addressing higher-order components:
 - "The AI didn't seem to understand my arguments in the way a human would."

≠ Human Feedback Perception:

- 85% of participants favored the personalized nature of human feedback, reflecting on its interplay between immediacy and personalization:
 - "The human feedback felt like a conversation; it motivated me to engage more with my writing."
 - "My teacher pointed out specific strengths and weaknesses that I hadn't considered before."
- Insights into efficiency reveal students appreciated immediate and thorough feedback:
 - "Having my instructor available to explain things in real-time was incredibly beneficial."
 - "Receiving extensive comments on my essays made me feel like I was getting a coaching session, which was much more effective than just automated replies."

4.2.2. Thematic Analysis

The data collected revealed several recurring themes that encapsulated participants' experiences:

Theme 1: AI-generated feedback - Immediacy and Limitations

≠ The immediacy of AI-generated feedback was often praised:

- “The AI was fast! I could see my mistakes immediately and correct them.”

≠ However, participants highlighted limitations in understanding deeper aspects of writing:

- “While the errors were highlighted promptly, I felt confused at times about how to improve my arguments.”

Theme 2: Human Feedback - Personalized Engagement

≠ The personalized nature of human feedback fostered deeper understanding:

- “My instructor took time to explain why certain phrases worked better; it felt more like coaching.”

≠ The engagement with instructors allowed students to address nuances effectively:

- “I really appreciated how my teacher pointed out specific areas to improve. It motivated me to take writing more seriously.”

Table 5

Summary of Qualitative Themes and Participant Quotes

| Theme | Participant Quote |
|--|---|
| AI-generated feedback Immediacy | - “The AI helped catch mistakes quickly, but didn’t guide me on how to fix my logic.” |
| Human Feedback Personalization | - “Receiving detailed comments was very encouraging—it felt like I could actually discuss my writing with someone.” |
| Value of Immediate Feedback | “AI tools made me aware of my errors quickly, but the human interaction clarified my doubts.” |
| Engaging with Feedback | “I could ask my instructor questions in real-time; with AI, I was often left to guess.” |

Participant Quotes:

1. “The AI was useful for quick corrections, but I think I really learned more from the depth of human feedback.”
2. “Being able to interact with my instructor helped me clarify my errors much better than just reading a list.”
3. “I preferred human feedback because it felt more tailored; the AI’s comments often seemed generic.”
4. “The corrections were immediate with AI, so I could adjust right away, but I often found the reasoning behind those corrections lacking.”
5. “Human feedback connected more with my learning style; I could ask questions right away.”

6. "I liked how my instructor pointed out patterns in my writing—something AI didn't do."
7. "While AI was quick, I sometimes had to guess its rationale, which was frustrating."
8. "The personal touch in human feedback made a difference; it felt like I was being mentored."
9. "I felt more encouraged with human feedback; it was clearer how to improve—not just what was wrong."
10. "AI-generated feedback was efficient, but I missed the human element; I needed someone to help with the 'why' behind my mistakes."

Through these insights, it becomes evident that students valued the immediacy offered by AI tools but simultaneously craved the nuanced understanding and motivational support that human feedback provides.

4.2.3. Summary of the Qualitative Findings

This study's qualitative findings provide in-depth insights into participants' perceptions of AI-generated and human feedback, revealing critical distinctions in their experiences. The responses, gathered through participant questionnaires, were analyzed thematically to explore how learners evaluated the effectiveness of each feedback type in improving their IELTS Writing performance.

Participants acknowledged the efficiency and consistency of AI-generated feedback, particularly in addressing surface-level errors such as grammar and vocabulary. The immediate nature of AI-generated feedback was highly valued, especially by learners managing tight schedules or preparing under time constraints. However, participants expressed dissatisfaction with the lack of interpretive depth, noting that AI often failed to address complex issues such as coherence, argument development, and contextual appropriateness. Some described AI-generated feedback as "generic" or "robotic," with limited adaptability to their individual needs.

Depth and Motivational Impact of Human Feedback

Human feedback was overwhelmingly preferred by participants for its personalized and interactive nature. Learners appreciated the detailed explanations and tailored suggestions provided by human instructors, which addressed not only their immediate errors but also the underlying causes. This type of feedback was particularly effective in enhancing task fulfillment, coherence, and argument structure. Many participants described human feedback as "engaging" and "motivational," highlighting its role in fostering confidence and encouraging them to take their writing seriously.

Comparative Analysis

The thematic analysis revealed clear distinctions in how learners perceived the two feedback types. AI-generated feedback was seen as a practical tool for immediate, surface-level corrections, while human feedback offered comprehensive support for both foundational and higher-order skills. Participants frequently expressed a preference for blended feedback approaches, combining the scalability and efficiency of AI with the depth and contextual adaptability of human instructors.

Key Themes from Participant Responses

The following themes emerged as central to the participants' experiences:

1. AI-generated feedback: Speed and Efficiency vs. Generic Suggestions

Strengths: Immediate error detection and correction in grammar and vocabulary.

Weaknesses: Limited depth in addressing coherence, argumentation, and contextual appropriateness.

2. Human Feedback: Tailored Guidance and Emotional Support

Strengths: Detailed explanations, adaptability to individual needs, and motivational impact.

Weaknesses: Time-intensive and less scalable compared to AI-generated feedback.

Generally, the qualitative findings reinforce the distinct yet complementary roles of AI and human feedback. While AI-generated feedback excels in efficiency and addressing technical errors, its inability to provide nuanced guidance highlights the enduring importance of human feedback for writing development. By combining the strengths of both, a blended feedback model can address learners' diverse needs and optimize their writing outcomes. While AI feedback was effective for immediate corrections, the qualitative data emphasizes that human feedback is essential for addressing higher-order skills, suggesting a need for a blended approach.

5. Discussion

This section interprets the findings by analyzing the strengths and limitations of AI-generated and human feedback on IELTS Writing performance. The results are contextualized within existing literature, highlighting implications for language pedagogy and high-stakes assessments like the IELTS.

The findings underscore the complementary roles of AI-generated and human feedback in addressing learners' writing needs.

Human feedback excelled in improving higher-order skills, particularly Task Achievement (TA), Coherence and Cohesion (CC), and Lexical Resource (LR). Participants noted its tailored nature, which allowed for contextualized corrections and deeper engagement.

Task Achievement and Motivation: Learners reported that personalized feedback helped clarify task requirements and fostered greater confidence. These findings align with Ferris (2006) and Nelson & Schunn (2009), who emphasize that scaffolding and personalized strategies are essential for addressing complex tasks. Regarding coherence and cohesion, human feedback provided actionable suggestions for improving logical flow and transitions, consistent with Hyland & Hyland (2006). For lexical resource, feedback on collocations and stylistic variation enhanced learners' ability to produce sophisticated vocabulary, supporting insights by Hyland (2003). The motivational impact of human feedback aligns with Lu and Law (2012), who argue that personalized, interactive feedback fosters learner confidence and engagement. Vygotsky's socio-cultural theory (1978) reinforces this view, highlighting how feedback acts as a scaffold for advancing learners' skills. AI-generated feedback was highly effective in addressing surface-level errors, particularly Grammatical Range and Accuracy (GRA) and repetitive lexical issues. Its efficiency and immediacy were particularly appreciated by learners preparing for IELTS under time constraints. Concerning grammatical range and accuracy, AI tools like ChatGPT effectively identified and corrected technical errors, resonating with Wilson & Roscoe (2020) and Ranalli et al. (2022). While AI-generated feedback provides efficiency and consistency, its limitations must be critically examined. One significant drawback is its difficulty in handling context-specific errors. Unlike human instructors, AI tools often fail to fully grasp a writer's intended meaning, leading to generic or misleading feedback on argument structure and coherence (Wang & Brown, 2020). For example, an AI system may flag a grammatically correct sentence as problematic due to an isolated structural anomaly, while overlooking deeper logical flaws in the overall argument.

Another challenge lies in idiomatic and stylistic accuracy. AI-generated feedback systems struggle with natural phrasing, collocations, and academic tone, often providing awkward or overly formal suggestions that do not align with IELTS writing expectations (Zhang & Hyland, 2020). Moreover, AI-generated feedback lacks the rhetorical awareness that human instructors bring, making it less effective in guiding students on genre-specific writing conventions.

To enhance the effectiveness of AI-generated feedback, several improvements can be considered. First, contextual learning algorithms could be integrated to enable AI tools to recognize holistic textual meaning rather

than isolated sentence-level errors. Additionally, AI systems could be trained on IELTS-specific writing corpora, ensuring that feedback aligns with the exam's unique linguistic and structural expectations. A promising direction is the development of hybrid feedback models, where AI provides initial corrections, followed by human verification for higher-order concerns like coherence and cohesion. Furthermore, adaptive AI systems that learn from individual users' writing patterns could offer personalized feedback, addressing recurrent weaknesses more effectively. By acknowledging these limitations and exploring potential solutions, future AI-generated feedback systems can move toward greater accuracy, contextual sensitivity, and pedagogical alignment with high-stakes writing assessments like the IELTS.

Despite these strengths, AI struggled to provide interpretive guidance for higher-order skills such as argumentation and coherence. These findings align with El Ebyary & Windeatt (2010) and Shirvani & Shirmohammadi (2020), who noted that AI-generated feedback often lacks the depth required for nuanced linguistic challenges.

The results advocate for a hybrid feedback approach, integrating the efficiency of AI with the depth of human feedback. Efficient Error Correction: AI-generated feedback can handle routine tasks like grammar correction, freeing human instructors to focus on higher-order concerns. This supports Ferris & Roberts (2001), who emphasized the importance of combining automated and human feedback for comprehensive skill development. Human feedback remains essential for guiding learners through complex writing processes, particularly in task-specific contexts. Blended models are especially valuable in large classrooms or online courses, where human feedback alone may not be feasible. This aligns with Elliot et al. (2021), who highlighted the scalability of AI tools, and Zhang & Hyland (2018), who advocate for hybrid systems to maximize learning outcomes.

This study highlights the complementary strengths of AI-generated and human feedback, advocating for hybrid systems that optimize writing instruction. While AI-generated feedback provides efficiency and consistency for surface-level errors, human feedback remains critical for higher-order skills and learner motivation. These insights offer actionable recommendations for educators, paving the way for innovative feedback mechanisms in language learning.

6. Conclusion

This study aimed to evaluate the comparative effectiveness of AI-generated and human feedback on the IELTS Writing performance of B2-level learners. The findings clearly highlight the strengths and limitations of each feedback type, emphasizing the need for a balanced, hybrid approach to maximize learning outcomes. By combining the efficiency of AI tools with the depth and personalization of human feedback, educators can create an optimal feedback system that addresses the diverse needs of learners preparing for high-stakes exams like the IELTS.

AI-generated feedback: Effective for addressing surface-level errors, such as grammar and vocabulary. Its immediacy and consistency make it a valuable resource for large classrooms, online settings, or time-constrained learning environments. However, its inability to address nuanced issues like coherence, cohesion, and argument development underscores the continued need for human intervention.

Human Feedback: Superior in improving higher-order skills such as task achievement, coherence, and lexical resource. The personalized nature of human feedback fosters motivation, builds confidence, and provides learners with actionable strategies for tackling complex writing challenges.

A significant finding of the study is the motivational impact of human feedback. Learners reported that personalized guidance increased their confidence and engagement with the writing process. AI-generated feedback also contributed to engagement by providing "quick wins," which encouraged learners to stay committed. These insights emphasize the importance of integrating feedback strategies that not only enhance technical skills but also support psychological and emotional aspects of learning. Beyond improving writing accuracy, this study highlights the motivational advantages of human feedback, as learners often feel more engaged and confident when receiving personalized guidance and encouragement from instructors. However, current AI-generated feedback systems lack

these motivational components, potentially limiting their effectiveness in fostering sustained learner engagement. To enhance AI-generated feedback, future systems could incorporate motivational elements similar to those found in human feedback. For instance, AI tools could use positive reinforcement strategies, framing feedback in an encouraging manner rather than focusing solely on corrections. Additionally, AI systems could integrate personalized progress tracking, providing learners with insights into their improvements over time, accompanied by motivational prompts (e.g., ‘Great job improving your coherence! Try adding more linking phrases for an even stronger argument.’). Another promising avenue is interactive feedback, where AI enables learners to ask for clarification or alternative suggestions, mimicking the dynamic engagement of human instruction. Furthermore, elements of gamification, such as achievement badges or personalized goal-setting, could help sustain learner motivation and commitment to writing improvement. By integrating these features, AI-generated feedback tools could become not only efficient writing assistants but also engaging learning companions, bridging the gap between automation and the human touch essential for effective language learning.

7. Pedagogical Implications

The findings of this study carry several important pedagogical implications for educators, institutions, and policymakers involved in language education. These implications center around the integration of AI-generated feedback into traditional teaching practices and the potential benefits of doing so:

1. **Blended Feedback Models:** Given that AI tools excel at providing rapid, corrective feedback on grammatical and lexical issues, while human feedback offers nuanced guidance on coherence and argumentation, a hybrid feedback model should be implemented in classrooms. Instructors should allow AI systems to handle lower-order concerns while reserving their own feedback for higher-order issues. This will enhance the efficiency of writing instruction, allowing more time to be spent on developing students' critical thinking and organizational skills.

2. **Time Management in Large Classrooms:** In large classes or online learning environments, it is often difficult for instructors to provide detailed feedback on every student's work. AI tools can serve as a valuable time-saving resource by identifying surface-level errors that do not necessarily require a teacher's attention. This allows teachers to focus their efforts on providing more personalized, in-depth feedback to students who need it most.

3. **Professional Development for Educators:** As AI tools become more integrated into educational settings, it is essential that educators receive training on how to use these tools effectively. Teacher training programs should include modules on how to interpret AI-generated feedback and how to use it as a starting point for more complex feedback discussions. Understanding AI's limitations will help teachers ensure that the technology is being used to supplement, not replace, human instruction.

4. **AI-generated feedback as Formative Assessment:** AI-generated feedback should be positioned as part of an ongoing formative assessment process. By providing immediate corrections on basic writing errors, AI tools allow students to see their mistakes in real-time and make revisions before submitting a final draft. This aligns with the concept of formative feedback, which is intended to help students improve their work as they progress through the learning process.

5. **Enhanced Feedback Delivery in IELTS Preparation**

Efficient Practice: AI-generated feedback can provide immediate corrections during practice sessions, helping learners internalize grammatical and lexical rules quickly.

Task-Specific Guidance: Human feedback is crucial for preparing learners to meet IELTS-specific requirements, such as adhering to task prompts and improving logical flow. Integrating both feedback types ensures comprehensive preparation for all aspects of the IELTS Writing criteria.

8. Limitations and Delimitations of the Study

While the study offers valuable contributions, certain limitations must be acknowledged:

The focus on short-term improvements leaves questions about the long-term effectiveness of the feedback types. The reliance on a single AI tool (ChatGPT) may limit the generalizability of findings to other systems. Cultural and contextual factors specific to Iran might influence the applicability of results in other regions.

Several delimitations were established to focus the scope of the study. First, the research was confined to B2-level learners preparing for the IELTS exam. This delimitation was intentional to ensure that all participants had a comparable level of language proficiency and writing ability. As a result, the findings may not extend to learners at different proficiency levels, such as C1 or C2, where writing expectations are significantly higher. Second, the study focused solely on writing performance. Other language skills, such as speaking, reading, and listening, were not considered, as the aim was to assess the specific impact of feedback on writing. Finally, the study limited its geographical scope to Iran, where specific cultural factors, educational practices, and attitudes toward technology might differ from other regions. This may influence how AI-generated feedback is received and utilized by learners. Expanding the research to other cultural contexts would offer more globally applicable insights.

9. Suggestions for Further Research

The findings of this study open several avenues for future research:

1. Exploring Different Proficiency Levels: Future research should explore the impact of AI-generated feedback and human feedback across different proficiency levels, from beginner (A1) to advanced (C2). This would help determine whether the effectiveness of each feedback type varies depending on the complexity of the writing tasks.
2. Longitudinal Studies: A longitudinal approach would be beneficial in assessing the long-term impact of AI-generated and human feedback on writing proficiency. Studies that track learner progress over several months or even years would provide valuable insights into the lasting effects of both feedback types.
3. Cross-Cultural Studies: Research could be expanded to include participants from different cultural backgrounds to examine how cultural factors influence perceptions and effectiveness of AI-generated feedback. This would help in understanding the global applicability of these tools and whether cultural attitudes towards technology play a role in feedback effectiveness.

References

- Almarwani, T., Alzahrani, M., & Al-Harbi, K. (2021). The effectiveness of AI-driven feedback tools in enhancing EFL learners' writing performance. *Journal of Educational Technology & Society*, 24(4), 23–34.
- Bitchener, J., Young, S., & Cameron, D. (2005). The effect of different types of corrective feedback on ESL student writing. *Journal of Second Language Writing*, 14(3), 191–205.
- Burgess, D., & Lacy, L. (2018). The role of AI in language education: From CALL to NLP and ITS. *Language Learning & Technology*, 22(3), 1–20.
- El Ebyary, K., & Windeatt, S. (2010). The role of feedback in second language writing. *Language Learning & Technology*, 14(1), 24–45.
- Elliot, A., Topping, K. J., & Hargreaves, D. J. (2021). Exploring blended learning in large classrooms: The role of AI tools in scalable feedback systems. *Educational Technology Research and Development*, 69(2), 405–425.
- Ferris, D. R. (2003). *Response to student writing: Implications for second language students*. London: Lawrence Erlbaum Associates.

- Ferris, D. R. (2006). Does error feedback help student writers? New evidence on the short- and long-term effects of different types of error feedback on ESL students' writing. *Journal of Second Language Writing*, 15(1), 1–18.
- Ferris, D. R. (2009). *Teaching ESL writing: Practical techniques in vocabulary and grammar*. Cambridge: Cambridge University Press.
- Ferris, D., & Roberts, B. (2001). Error feedback in L2 writing: How explicit does it need to be? *Journal of Second Language Writing*, 10(4), 261–300.
- Grabe, W., & Kaplan, R. B. (1996). *Theory and practice of writing: An applied linguistic perspective*. London: Longman.
- Green, A. (2007). *Research on IELTS: A report on the research commissioned by the IELTS partners*. Cambridge: Cambridge University Press.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112.
- Hyland, K. (2003). *Second language writing*. Cambridge: Cambridge University Press.
- Hyland, K., & Hyland, F. (2006). Feedback on second language students' writing. *Language Teaching*, 39(2), 83–101.
- Hyland, K., & Hyland, F. (2019). *Feedback in second language writing: Contexts and issues*. Cambridge: Cambridge University Press.
- IELTS. (2021). *IELTS scoring in detail*. ETS: IELTS Official.
- Lu, X., & Law, F. (2012). A study on the effectiveness of teacher and peer feedback on ESL writing. *Language Learning & Technology*, 16(1), 44–61.
- Lu, X., & Law, N. (2012). A meta-analysis of studies on the effectiveness of computer-assisted feedback in language learning. *Language Learning & Technology*, 16(3), 16–32.
- Maleki, A., & Zare, P. (2015). The effect of IELTS preparatory courses on Iranian EFL learners' writing performance. *Journal of Language Teaching and Research*, 6(3), 581–587.
- Moinzadeh, A., & Zafarghandi, A. M. (2015). The role of critical thinking in Iranian EFL learners' academic writing. *Language Teaching Research*, 19(3), 315–329.
- Nelson, S., & Schunn, C. D. (2009). Scaffolding feedback for peer review in an online learning environment. *Educational Psychology*, 29(1), 41–59.
- O'Sullivan, B., & Green, A. (2003). *Cambridge IELTS 3: Examination papers from the University of Cambridge ESOL examinations*. Cambridge: Cambridge University Press.
- O'Sullivan, B., & Weir, C. (2011). Investigating the role of automated feedback in language learning: A study of the IELTS Writing test. *Language Testing*, 28(4), 555–577.
- Piaget, J. (1970). *Science of education and the psychology of the child* (D. Coltman, Trans.). Viking Press. (Original work published 1969).
- Rahimi, M., & Kafipour, R. (2014). Coherence in Iranian EFL learners' essays: Challenges and strategies. *TESL-EJ*, 18(1), 1–14.

- Ranalli, J., Sinha, M., & Southgate, L. (2022). The role of AI in second language writing: Perspectives from research and practice. *Journal of Writing Research, 14*(2), 301–318.
- Shirvani, H., & Shirmohammadi, M. (2020). AI-generated feedback systems in language learning: Impact on writing performance in EFL classrooms. *Computers & Education, 14*(6), 445–467.
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153–189.
- Skinner, B. F. (1954). The science of learning and the art of teaching. *Harvard Educational Review, 24*(2), 86–97.
- Tavakoli, H., & Pashmforoosh, R. (2015). Challenges of adopting Western rhetorical styles: Perspectives from Iranian EFL learners. *Journal of English for Academic Purposes, 21*(1), 23–34.
- Taylor, L., & Weir, C. (2012). *IELTS: A user's guide*. London: Routledge.
- Vahdat, S., & Khodabakhsh, S. (2019). The impact of peer and teacher feedback on Iranian EFL learners' writing quality. *Language Testing in Asia, 9*(2), 1–12.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Wang, X., & Brown, C. (2020). Comparing AI and teacher feedback in an ESL writing course: Effects on grammar, coherence, and overall writing skills. *Journal of Second Language Writing, 48*, 100–116.
- Ware, P., & Warschauer, M. (2006). Comparing digital feedback tools in second language writing. *Language Learning & Technology, 10*(3), 1–20.
- Warschauer, M., & Grimes, D. (2008). Automated feedback and human feedback in writing instruction. *Language Learning & Technology, 12*(3), 63–79.
- Wilson, A. M., & Roscoe, R. D. (2020). AI in educational feedback: A review of applications in writing and language learning. *Computers & Education, 14*(8), 312–328.
- Xie, Y., Chu, H. C., Hwang, G. J., & Wang, H. P. (2019). The effects of AI-based and human feedback in a blended learning environment: A comparative study. *Computers & Education, 14*(2), 103–123.
- Zarei, A. A., & Hashemnezhad, H. (2020). The role of AI tools in IELTS preparation: Evidence from Iranian EFL learners. *Journal of Language Teaching and Research, 11*(6), 1023–1030.
- Zarei, G. R., & Sadighi, F. (2011). Challenges of IELTS writing tasks: Insights from Iranian learners. *Iranian Journal of Applied Linguistics, 14*(2), 71–89.
- Zhang, S., & Hyland, K. (2020). Learner perceptions of AI-generated feedback in EFL writing: Benefits, limitations, and implications. *Journal of English for Academic Purposes, 44*, 108–123.
- Zhang, Y., & Hyland, K. (2018). Hybrid feedback systems: Combining teacher and AI-generated feedback in EFL contexts. *Journal of Writing Research, 10*(1), 56–78.

