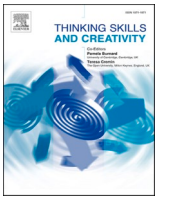
Thinking Skills and Creativity 56 (2025) 101766

Contents lists available at ScienceDirect

Thinking Skills and Creativity

journal homepage: www.elsevier.com/locate/tsc



The impact of integrating ChatGPT with teachers' feedback on EFL writing skills

Marjan Asadi [a,1], Saman Ebadi [b,2,*], Laleh Mohammadi [c,3]

[a] Department of English Language and Literature, Faculty of Literature and Humanities, Tehran University, Tehran, Iran
[b] Department of English Language and Literature, Faculty of Literature and Humanities, Razi University, Kermanshah, Iran
[c] Islamic Azad University, Isfahan, Iran

ARTICLE INFO

Keywords:
ChatGPT
AI
Adaptive feedback
Personalized writing instruction
Teacher feedback
Ethical AI integration

ABSTRACT

This study examined the effects of integrating ChatGPT, a state-of-the-art AI chatbot, with teacher feedback to provide individualized writing feedback on English as a foreign language (EFL) essay writing task 2. It also explored the perspectives of eight teachers on integrating such AI tools ethically and effectively. The study employed an explanatory sequential mixed-methods design and randomly assigned 68 intermediate Iranian EFL learners to receive either a combination of ChatGPT and teacher feedback or teacher-only feedback on their IELTS Task 2 argumentative essays. The results of the Paired Samples t-Test indicated that the group that received a combination of teacher and ChatGPT feedback demonstrated a significantly greater improvement than the group receiving only teacher feedback on their IELTS Task 2 argumentative essays in all scoring criteria, including task achievement, coherence, cohesion, vocabulary, and grammar range and accuracy. Thematic analysis of semi-structured interviews with teachers in both conditions revealed concerns about plagiarism and overreliance on AI. The findings of this study highlight the transformative potential of integrating AI tools with traditional teaching methods to deliver personalized feedback, significantly advancing EFL learners' writing proficiency and addressing practical challenges in writing instruction.

1. Introduction

Developing writing skills proficiency remains an ongoing challenge for many English language learners (Wei, Wang & Dong, 2023). According to Cao and Zhong (2023), feedback is an essential component of ESL/EFL writing instruction, but it can have different effects depending on the type of feedback. Previous studies indicated timely and effective feedback is essential for improving their writing skills (Su, Lin & Lai, 2023). Despite traditional classroom instruction and structured teacher feedback, students often lack the personalized guidance to enhance their writing abilities (Lee, Liang & Yang, 2022). The rapid development of educational technology

* Corresponding author.
E-mail address: samanebadi@gmail.com (S. Ebadi).

[1] Marjan Asadi holds an MA in Applied Linguistics from Tehran university. She is interested in AI contributions in education.

[2] Saman Ebadi is a full professor in Applied Linguistics at Razi University, Kermanshah, Iran. His main research interests include Mobile Assisted Language Learning, Dynamic Assessment, and Blended Learning. He has published several papers in internationally acclaimed journals like Computer Assisted Language Learning, Education and Information Technologies and CALL-EJ.

[3] Laleh Mohammadi is a BA in Public Management. She is interested in AI in education.

https://doi.org/10.1016/j.tsc.2025.101766
Received 25 October 2024; Received in revised form 16 January 2025; Accepted 16 January 2025
Available online 23 January 2025

has created opportunities to integrate technology into classroom instruction (Hyland & Hyland, 2006). Similarly, Yüce (2019) emphasized that while online foreign language teaching has been largely well-received by instructors in terms of improving language skills and areas, notable challenges persist, especially concerning classroom management. This highlights the ongoing need for effective strategies to balance the benefits of technology with the demands of maintaining an engaging and organized learning environment. The use of technology for teaching and learning language skills, especially writing, has seen significant growth in recent years (Godwin-Jones, 2019). The technology used to improve writing quality and the learners' behavior has been highlighted by many studies (e.g., Potter & Wilson, 2021; Waer, 2023; Zhu, Liu & Lee, 2020). Similarly, Yüce (2022) emphasized the role of digital tools in language education during the COVID-19 pandemic, noting that while online platforms facilitated language learning, they also brought challenges, particularly in maintaining engagement and effective classroom management. This growing reliance on technology underscores the need for a balanced approach to integrating digital tools into language learning environments.

Artificial Intelligence (AI) aims to create machines that can emulate human intelligence processes (Dogan, Goru Dogan & Bozkurt, 2023). Karataş and Yüce (2024) stated that the rapid advancement of AI in education presents transformative opportunities for open and distributed learning environments, including distance, hybrid, and blended settings. This highlights AI's potential to enhance educational effectiveness while also raising concerns about ethical considerations and data security. One kind of AI, called Large Language Models (LLM), can generate text that resembles human writing. A common application of LLM is a 'chatbot,' a software program that can simulate human conversation, such as ChatGPT, Bard, and Bing. ChatGPT, a state-of-the-art natural language processing model, demonstrates strong language comprehension and generation abilities (Liu et al., 2023). It could be used to evaluate assignments and give feedback to students in real-time, enabling a more effective and personalized learning experience (J. Gao, 2021; Roscoe, Allen, Johnson & McNamara, 2018; Zawacki-Richter, Marín, Bond & Gouverneur, 2019). It can automatically grade essays with reasoning and suggestions and assist with writing speeches by generating ideas, summaries, and even full scripts (Javaid, Haleem, & Singh, 2023). This integration of AI presents opportunities for personalized learning but requires careful attention to ethical concerns and teacher training.

ChatGPT offers a multifaceted approach to enhance writing skills. It enables teachers to design prompts for open-ended questions aligned with instructional objectives and success criteria (Baidoo-Anu & Owusu Ansah, 2023). In written communication, ChatGPT excels at producing clear and concise language, assisting learners in honing their language skills (Javaid et al., 2023). Beyond that, it demonstrates an ability to engage in back-and-forth discussions, encouraging interactive writing experiences (Alkaissi & McFarlane, 2023). This multifunctional tool showcases its versatility in supporting writing development and language proficiency. ChatGPT offers the potential to provide individualized and interactive help tailored to each learner's unique preferences and requirements. This individualized and interactive support enhances learning experiences, fostering learners' autonomy, engagement, and motivation (Firat, 2023). The capabilities of ChatGPT have been confirmed through various assessments and examinations (Dwivedi et al., 2023; Li, Wei & Lu, 2023). ChatGPT's capabilities assist in language development and foster learners' autonomy, engagement, and motivation through tailored, interactive support, significantly enhancing the writing process.

This study adopts a sociocultural theory framework to better understand the integration of AI tools like ChatGPT and traditional teacher feedback in enhancing EFL learners' writing skills. From Vygotsky's perspective, sociocultural theory emphasizes the importance of social interaction, cultural tools, and mediated learning in cognitive development. AI tools, such as ChatGPT, can be viewed as mediators within the Zone of Proximal Development (ZPD), where learners develop skills they could not achieve independently through guided support ((Gauvain, 2020; Marginson & Dang, 2017)). Combining real-time, personalized feedback from AI with the contextual expertise of teacher feedback fosters a collaborative and scaffolded learning environment that enhances students' writing proficiency (Shabani, 2016). This theoretical lens underscores how technology can serve as a transformative tool to support autonomous learning and teacher-student interactions, addressing the ongoing challenge of developing writing skills proficiency among English language learners (Panofsky, 2003).

The emergence of adaptive learning technologies, such as ChatGPT, holds promise for enhancing education. However, effective implementation of these technologies requires the establishment of elusive best practices (Lu, Qiu, Ding, Xie & Tao, 2023). Previous research has criticized the efficacy of technology-based writing assessment tools (e.g., Fu, Zou, Xie, & Cheng, 2022; Herrington & Moran, 2012; Lang, Aull & Marcellino, 2019). Studies on the effects of various feedback modes on writing performance (e.g., teacher, automated, and blended) have yielded inconsistent results (e.g., Cunningham & Link, 2021; Wilson & Czik, 2016; Zhang & Hyland, 2018). While existing studies have explored the individual effects of feedback modes, there remains a gap in understanding the synergistic impact of combining AI tools like ChatGPT with traditional teacher feedback in improving writing performance. This study aims to fill this gap by examining how integrating AI and teacher feedback can enhance EFL learners' writing skills, offering practical insights for educators and policymakers.

A notable research gap exists in understanding how different feedback types influence students' writing progress and teachers' perceptions of these modes (Mohammadi, Zarrabi, & Kamali, 2023). While prior studies have examined the individual impacts of teacher and AI-driven feedback, their combined effectiveness in addressing students' specific needs remains underexplored. This integration is particularly significant in L2 writing, where combining AI's personalized, real-time suggestions with teachers' contextual expertise can create a more effective, adaptive, and student-centered feedback framework. Investigating this synergy is critical, as it can reveal how AI and teacher feedback can complement each other to optimize learning outcomes.

Furthermore, understanding teachers' perceptions of AI integration is crucial because these perceptions directly influence how effectively AI tools like ChatGPT are utilized in practice. Teachers, guided by their professional judgment and knowledge of students' specific writing challenges, play a key role in determining how ChatGPT's personalized feedback can be tailored to meet individual learner needs. By examining the influential factors shaping teachers' evaluations, such as usability, reliability, and alignment with pedagogical goals, this study provides insights into how AI can be effectively integrated to enhance L2 writing instruction.

Additionally, the research explores whether interactions with ChatGPT necessitate scaffolding or supplementary human guidance to fully unlock its potential for fostering language development (Kohnke, Moorhouse, & Zou, 2023).

The following research questions guided the study:

RQ1: How does adaptive ChatGPT feedback influence EFL learners' writing skills compared to traditional teacher feedback, and what specific aspects of writing are most affected?

RQ2: In what ways do EFL teachers perceive the effectiveness of AI-assisted feedback in enhancing EFL writing skills, and what factors contribute to their evaluations?

2. Literature review

2.1. Teacher writing feedback

Providing feedback is an essential part of the learning process, particularly in writing instruction, as it directly impacts students' learning outcomes and skill development. Teacher feedback is a significant factor in improving writing skills, as it offers personalized, contextualized guidance that aligns with individual students' needs (Ferris, 2011; Hyland & Hyland, 2006). Research has shown that effective teacher feedback addresses surface-level errors and nurtures deeper understanding, critical thinking, and the ability to self-reflect (Hattie & Timperley, 2007). For instance, teacher feedback can enhance students' metacognitive strategies, including planning, monitoring, and evaluating their writing, leading to improved writing proficiency (Zhang & Zhang, 2024). Furthermore, personalized feedback from teachers is often perceived as more valuable and effective than other feedback types, including peer or automated feedback, because it provides tailored insights that directly address individual students' challenges (Cunningham, 2019; Liu & Wu, 2019). However, this study aims to explore a gap in the literature regarding integrating AI feedback with teacher feedback.

However, despite its strengths, providing timely and personalized feedback poses significant challenges, particularly in online or large classroom settings. Teachers may struggle with workload constraints, limited time, and difficulty managing individualized feedback for many students. These challenges are amplified in online learning environments, where the lack of face-to-face interaction can hinder nuanced communication and the establishment of rapport between teachers and students (Xu, Meng, Raja, Priya & Kiruthiga Devi, 2023). Consequently, there is a growing need to support teachers in these settings by leveraging additional resources, such as AI-based tools, that can complement their efforts. Feedback from educators' training can also serve as a vital resource to equip teachers with strategies and tools to overcome these challenges. Such training emphasizes the integration of AI technologies to generate tailored feedback that aligns with students' specific needs and preferences, easing the burden on teachers while maintaining high-quality feedback delivery.

2.2. AI-Based writing feedback studies

AI-driven feedback systems, powered by advancements in natural language processing, have emerged as transformative tools in writing instruction. These systems analyze students' writing and provide immediate, consistent, and personalized feedback, addressing grammar, syntax, content, and overall structure (Clarizia, Colace, Lombardi, Pascale & Santaniello, 2018; Taskiran & Goksel, 2022). For instance, ChatGPT and similar tools use generative AI technologies to offer instant, adaptive feedback that exposes learners to authentic language usage and improves their writing fluency and accuracy (Hong, 2023; Kim, Shim & Shim, 2023). The speed and scalability of AI feedback systems make them particularly advantageous in addressing the challenges faced in large or online learning environments, where timely and individualized feedback may otherwise be unattainable (Xu et al., 2023). Therefore, more studies are needed to explore the effectiveness of integrating AI-driven feedback with traditional methods to enhance writing outcomes and identify best practices for various educational contexts.

Studies have demonstrated the potential of AI tools to improve student outcomes. For example, Su et al. (2023) found that ChatGPT effectively enhances EFL learners' argumentative writing by providing detailed feedback on grammar, structure, and content while also boosting students' motivation and self-efficacy. Similarly, Taskiran and Goksel (2022) observed that AI-driven feedback tools, such as Pigaiwang software, significantly improve academic writing skills in open and distance learning contexts. However, the effectiveness of AI feedback is influenced by factors such as the quality of the training data, the complexity of the task, and students' proficiency levels. While AI tools excel at addressing surface-level issues, they cannot often provide deeper, context-specific insights, which are typically offered by human teachers (Grassini, 2023; Wang & Demszky, 2023). Therefore, future research should focus on optimizing the integration of AI tools with human feedback to bridge this gap and enhance the overall learning experience.

Researchers have increasingly highlighted the benefits of integrating AI feedback with teacher feedback to address this limitation. A combined approach leverages the strengths of both, with AI handling immediate and repetitive tasks while teachers provide more nuanced, contextualized feedback that fosters critical thinking and deeper engagement with the writing process (Asadi & Taheri, 2024; Guo & Wang, 2023). For instance, Asadi and Taheri (2024) noted that combining AI-generated feedback with teacher guidance significantly enhances students' learning experiences, fostering collaboration and improving outcomes in high-stakes assessments like the IELTS.

Despite these advancements, critical gaps remain in understanding how teachers perceive the integration of AI feedback into L2 writing instruction and the factors shaping these perceptions. Teachers' acceptance and evaluation of AI tools are pivotal, as their perceptions directly influence how effectively tools like ChatGPT are implemented in practice and their impact on students' learning outcomes (Al-Garaady & Mahyoob, 2023). Teachers, guided by their professional expertise and understanding of students' specific

writing needs, play a central role in tailoring AI-generated feedback to individual learners. This study addresses these gaps by investigating how EFL teachers evaluate ChatGPT's feedback about students' writing skill development. It also examines the influential factors that shape these evaluations, such as usability, reliability, and alignment with pedagogical goals.

Furthermore, this study aims to explore how adaptive AI feedback, when integrated with teacher feedback, influences EFL learners' writing skills. By focusing on the comparative effectiveness of ChatGPT and traditional teacher feedback, this research will identify which aspects of writing are most impacted by this synergy. The findings will offer insights into how AI and teacher feedback can complement each other to optimize EFL writing outcomes.

3. Methodology

3.1. Design

The study employed a sequential explanatory mixed-methods design, consisting of a quantitative quasi-experimental phase followed by a qualitative phase (Creswell, Clark, Gutmann & Hanson, 2003). The quantitative phase involved pre-and post-test assessments with two groups: an experimental group that received both ChatGPT-assisted and traditional teacher feedback and a control group that received only traditional feedback. This design allowed for a direct comparison between the two approaches to investigate their impact on students' argumentative writing skills.

The qualitative phase involved semi-structured interviews with eight teachers, exploring their perceptions of the benefits and drawbacks of ChatGPT-assisted feedback compared to traditional feedback methods. These interviews were analyzed using thematic analysis (Braun & Clarke, 2006).

3.2. Participants and selection

The study included 68 intermediate English language learners (aged 21–41; 43 females, 25 males) preparing for the IELTS Writing Task 2. Participants were enrolled in a 12-session IELTS writing preparation course at a language institute in Iran. The institute's placement test assessed their proficiency, which evaluated writing, grammar, and vocabulary. Learners scoring within the B1–B2 range on the Common European Framework of Reference (CEFR) were selected, ensuring a homogenous group of intermediate English learners. An additional IELTS mock test confirmed participants' writing proficiency, with most achieving band scores between 5.5 and 6.0 on Writing Task 2.

Purposeful sampling was employed to select learners who were considered to benefit most from the writing interventions based on their proficiency level and course enrollment (Creswell & Creswell, 2017). Intermediate learners were chosen based on the assumption that they would benefit most from targeted writing interventions. This group is proficient in engaging with complex tasks but still requires improvement to achieve higher IELTS band scores (7.0 or above). Candidates scoring below B1 were excluded, resulting in a final cohort of 68 participants. All participants were fully informed about the study's purpose, provided written consent, and were assured of their rights throughout the process.

Participants were randomly assigned to either the experimental group, which received feedback from ChatGPT and teachers or the control group, which received teacher-only feedback. This randomized assignment ensured a balanced comparison of the two feedback conditions.

The study also included eight IELTS-certified teachers from the same language institute. These teachers were certified IELTS trainers and former examiners, with expired certificates demonstrating their expertise in IELTS standards and candidate needs. All instructors underwent training on effectively integrating ChatGPT into their feedback processes to standardize their familiarity with AI tools. This training emphasized understanding the strengths and limitations of AI-generated feedback, customizing prompts, and enhancing AI-generated suggestions with pedagogical insights to align with student needs.

To further ensure consistency in the integration of ChatGPT feedback, a shared framework for making queries to the AI was implemented. This framework provided uniform guidelines for structuring prompts, ensuring that all teachers used ChatGPT in a standardized manner to generate relevant and accurate feedback for student essays. Developed during the training sessions, the framework included specific templates and examples to guide teachers in making effective queries. These measures minimized variability in feedback quality across the experimental group, enhancing the study's methodological rigor.

To minimize bias and ensure objectivity, the instructors who taught the six-week course and provided feedback on students' essays were distinct from those who conducted pre- and post-test assessments. This separation prevented potential biases by ensuring that post-test evaluations were conducted blind, with assessors unaware of group assignments. Such measures mitigated risks of confirmation bias and safeguarded the objectivity of the study.

Ethical considerations were rigorously addressed to ensure the rights and well-being of all participants. The selection process prioritized transparency and informed consent. Students were thoroughly briefed on the study's objectives, procedures, and voluntary nature, and written consent was obtained from all participants. Teachers were similarly informed about the study's purpose and assured of their voluntary participation, with guarantees of confidentiality regarding their input and feedback. Adhering to ethical guidelines established by Iranian language institutes, the study ensured compliance with professional standards, safeguarding participants' rights while maintaining academic integrity throughout the research process.

3.3. Instruments

3.3.1. IELTS academic writing task 2

Task 2 of the IELTS Academic Writing section was used for both pre- and post-testing. Writing tasks during the instructional phase were selected from The Key to IELTS Writing by Cullen (2020), providing a consistent and comprehensive framework for the experimental and control groups. This book was chosen for its practical approach to teaching strategies and skills essential for writing clear, logical, and coherent essays in both academic and general training formats. It addresses common challenges learners face, aiming for band scores of 7 or higher, including generating ideas, planning essay outlines, crafting introductions and conclusions, using appropriate vocabulary and grammar, and proofreading effectively.

The book adopts a step-by-step methodology for writing essays, guiding students from analyzing the prompt and forming a concise argument to developing cohesive ideas supported by examples. The writing tasks are also based on authentic and reliable sources, such as IELTS Cambridge 9 test questions, ensuring their relevance to the official IELTS exam.

Sample prompts used during the study included:

- Some experts believe children should begin learning a foreign language at primary rather than secondary school. Do the advantages of this outweigh the disadvantages?
- Some believe unpaid community service should be compulsory in high school programs (e.g., working for a charity, improving the neighborhood, or teaching sports to younger children).
- Some people say the best way to improve public health is by increasing the number of sports facilities. However, others believe this would not affect public health and that other measures are required. Discuss both these views and give your own opinion.

3.3.2. Semi-structured interviews

A semi-structured interview protocol was developed based on established qualitative research standards and piloted with a professional IELTS teacher to ensure the appropriateness and clarity of the questions. Feedback from the pilot interview was incorporated to refine the protocol further. Semi-structured interviews were subsequently conducted with eight IELTS-certified teachers to explore their perspectives on integrating AI-assisted feedback in IELTS writing preparation. The interviews were audio-recorded and transcribed verbatim, and the data were analyzed using thematic analysis, as outlined by Braun and Clarke (2006). This method enabled the identification of recurring themes and nuanced insights into the teachers' experiences.

Strategies recommended by Lincoln and Guba (1985) were employed to enhance the credibility and trustworthiness of the findings, including peer debriefing and member checking. These approaches ensured the reliability of the analysis and minimized researcher bias. For a full list of interview questions, see Appendix 1.

3.3.3. Procedure

A 12-session IELTS writing preparation course was designed to enhance participants' argumentative writing skills and help them achieve a target band score of 7.0. The course was based on The Key to IELTS Writing Task 2 (Cullen, 2020), which provided a structured approach to essay writing, including essential skills for writing clear, logical, and coherent arguments. Tasks and models from the book, which align with authentic IELTS Cambridge 9 test prompts, were utilized to ensure relevance to the exam.

Participants were randomly assigned to an experimental group (EG, $n = 34$) or a control group (CG, $n = 34$) after completing the pre-test to ensure balanced proficiency levels across groups. Randomization was conducted using a computer-generated sequence based on placement test scores and mock IELTS Writing Task 2 results.

Both groups attended 12 sessions, each lasting 90 min, focusing on argumentative essay structures, cohesive writing strategies, and language use in academic contexts. In the first session, all participants completed a pre-test, writing an IELTS Task 2 essay in 40 min, which was scored by teachers using the IELTS Writing Band Descriptors (task achievement, coherence and cohesion, lexical resource, and grammatical range and accuracy). Calibration sessions among teachers ensured scoring consistency. For both groups, explicit instruction was provided on high-scoring argumentative writing techniques, including generating ideas, structuring essays, and applying academic vocabulary. After receiving instruction, students submitted two drafts per essay for feedback. The first draft was submitted after initial instruction, and the second draft was submitted after receiving detailed feedback. This process allowed students to revise and improve their essays iteratively.

The six essays written by students during the course were not merely repeated versions of the pre-course writing task. Instead, the pre-course writing task was used as a diagnostic tool to assess students' baseline performance across the IELTS Writing Task 2 band descriptors (task achievement, coherence and cohesion, lexical resource, and grammatical range and accuracy). This provided teachers and students with a clear understanding of their needs for improvement. One of the six essays assigned during the course was identical to the pre-course writing task, allowing students to directly apply the skills they had developed through feedback and targeted practice.

Each of the six essays focused on a distinct aspect of argumentative writing, such as generating ideas, planning, introduction and conclusion writing, improving coherence and cohesion, academic vocabulary, and grammar accuracy. For each essay, students submitted two drafts: the first draft was followed by detailed feedback, and the second draft allowed for further revisions based on this feedback.

- Experimental Group (EG):

The EG received combined feedback from teachers and ChatGPT. Teachers crafted prompts to elicit feedback from ChatGPT on

various aspects of writing, including grammar, vocabulary, coherence, and task achievement. Teachers reviewed, validated, and supplemented AI-generated feedback, ensuring high-quality, personalized guidance. For instance, teachers used ChatGPT to suggest advanced vocabulary or refine underdeveloped ideas, enriching this feedback with their pedagogical insights. ChatGPT feedback was incorporated into teacher comments, fostering student confidence and highlighting AI as a learning tool. A shared framework for making queries to ChatGPT was implemented to ensure consistency in feedback, which provided uniform guidelines for structuring prompts across all teachers.

- **Control Group (CG):**

The CG received feedback from the traditional teacher on the same writing elements. Teachers focused on vocabulary, grammar, content, and organization, providing personalized comments for improvement. Feedback was returned within one week to guide revisions and essay finalization before subsequent submissions.

At the end of the course, all participants completed a post-test, where essays were scored using the same IELTS Writing Band Descriptors. Final scores were derived through consensus among raters to ensure accuracy. For additional details on feedback examples, see Appendices 2, 3, and 5.

3.4. Data analysis

Quantitative data were analyzed using SPSS version 25 (Statistical Package for the Social Sciences). A paired-sample t-Test was performed to examine improvements in the writing performance of the experimental group (EG) before and after the treatment. This test was appropriate as it measures the mean differences within the same group across two-time points. An independent-sample t-Test was conducted to compare the writing performance of the EG and control group (CG) after the treatment. This statistical test was chosen to assess the mean differences between two independent groups (Pallant, 2020).

Qualitative data from the teacher interviews were analyzed using thematic analysis, following the guidelines outlined by Braun and Clarke (2006). Thematic analysis was employed as it provides a systematic approach to identifying, organizing, and interpreting patterns of meaning within qualitative data. The analysis involved four steps:

- Transcribing the audio recordings of the interviews.
- Examining the transcripts for meaningful features and generating initial codes.
- Categorizing the codes into themes and sub-themes.
- Reviewing and refining themes to ensure coherence and eliminate redundancy or overlap.

This mixed-methods approach ensured a comprehensive analysis of quantitative and qualitative data, enhancing the study's validity and depth of insights.

4. Results

4.1. Addressing the first research question

The first research question investigated whether a combination of ChatGPT and teacher feedback improved English learning outcomes more than teacher feedback alone. To address this research question, initially, the results of the pretest for the experimental and control groups in terms of task response (TR), cohesion and coherence (CC), lexical resource (LR), grammatical accuracy (GA), and overall score (OS) were compared using a series of independent samples t-Tests to make sure that the two groups were not significantly different in regard to these variables at the outset of the study. Table 1 displays the descriptive statistics of the experimental and control groups for the TR, CC, LR, GA, and OS pretest results.

As shown in the table above, the Skewness and Kurtosis ratios for the pretest scores all fell within the $\pm$ 1.96 range. Therefore, since the normality assumption was not violated, it was possible to run the parametric test of independent samples t-Test (Tabachnick &

Table 1
Descriptive statistics of the experimental and control groups for the TR, CC, LR, GA, and OS pretest results.

	Group	N	Mean	Std. Deviation	Std. Error Mean	Skewness Ratio	Kurtosis Ratio
TR.pr	Experimental Group	34	2.41	.74	.12	−1.12	1.38
	Control Group	34	2.35	.77	.13	1.44	−1.56
CC.pr	Experimental Group	34	2.00	.69	.11	0.88	0.36
	Control Group	34	1.82	.67	.11	1.54	0.04
LR.pr	Experimental Group	34	2.38	.49	.08	0.07	1.07
	Control Group	34	2.35	.48	.08	−1.25	1.33
GA.pr	Experimental Group	34	2.55	.50	.08	0.11	−1.47
	Control Group	34	2.50	.66	.11	0.84	0.17
OS.pr	Experimental Group	34	9.38	1.25	.21	1.32	0.89
	Control Group	34	9.02	1.46	.25	0.81	−1.65

Fidell, 2007). Table 2 displays the results of independent samples *t*-tests with Welch correction for the experimental and control pretest score means.

As indicated in the above table, all the sig indices (two-tailed) are above the critical value of 0.05. Accordingly, it can be inferred that the two groups were not significantly different in terms of TR, CC, LR, GA, and OS prior to the treatment. Thus, any differences between the score mean of the two groups on the posttest can be attributed to treatment. Table 3 demonstrates the descriptive statistics of the experimental and control groups for the TR, CC, LR, GA, and OS posttest results.

As shown in the above table, the Skewness and Kurtosis ratios for the post-test scores fell within the ± 1.96 range. Therefore, because the normality assumption was met, running the parametric test of independent samples t-Test was warranted (Tabachnick & Fidell, 2007). Table 4 presents the results of independent samples t-Tests with Welch correction for the experimental and control post-test score means.

As presented in the above table, all the sig indices (two-tailed) are below 0.001. Accordingly, it can be inferred that the two groups were significantly different regarding TR, CC, LR, GA, and OS after the treatment. Thus, it can be concluded that the implementation of treatment has significantly affected the participants' learning outcomes with respect to TR, CC, LR, GA, and OS. To calculate the effect sizes, that is to determine the strength of the findings and evaluate the stability of the research findings across samples, Cohen d was calculated using formula (#1) in Lakens (2013, p. 3).

$$d_s = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{(n_1-1)SD_1^2+(n_2-1)SD_2^2}{n_1+n_2-2}}} \tag{1}$$

The calculated effect sizes for the TR, CC, LR, GA, and OS turned out to be 1.33, 1.90, 1.24, 1.28, and 3.00, respectively, which are all considered large effect sizes according to Cohen (1988). Thus, the findings can be safely generalized.

The pretest and posttest results of the experimental group were compared to provide further assurance of the positive influence of treatment on learning outcomes. Since the data sets belonging to the experimental group's pretest and posttest were normally distributed (See Tables 1 and 3), the experimental group's pretest and posttest score means were analyzed using a series of paired samples t-Tests. Table 5 portrays the respective results.

As revealed in the above table, all the sig indices (two-tailed) turned out to be lower than 0.001. Therefore, it can be inferred that the experimental group scored higher on posttests than pretests in terms of TR, CC, LR, GA, and OS. Such results can provide further assurance of the positive influence of treatment on learning outcomes. The effect sizes were computed using Cohen's d_z formula (#7) proposed by Rosenthal (1991, cited in Lakens, 2013, p. 4) as follows:

$$\text{Cohen}'s d_z = \frac{\text{t}}{\sqrt{\text{n}}} \tag{2}$$

The computed effect sizes for the TR, CC, LR, GA, and OS equaled 1.68, 1.91, 1.59, 1.54, 2.92, respectively, which are all regarded large (Cohen, 1988). Accordingly, such results have practical significance and can be safely generalized across samples.

4.2. Thematic analysis of interview data

This study employed thematic analysis to examine transcribed interview data and identify key themes systematically. Thematic analysis is a qualitative research method that enables researchers to explore and interpret patterns of meaning within the data (Braun & Clarke, 2006). A flexible set of guidelines was followed to ensure rigor and consistency in analyzing the data and developing data-driven themes. Using open coding principles, the researchers identified core variables, grouped related codes into clusters, and labeled each cluster based on its shared theme.

A bottom-up (inductive) approach allowed themes to emerge organically from the data without preconceived assumptions. To ensure reliability, two independent coders reviewed the transcripts and identified themes. The inter-rater reliability was assessed using Cohen's Kappa coefficient, yielding a value of 0.82, which reflects a strong agreement between the coders. Additionally, an impartial reviewer assessed the coherence of the themes to ensure their alignment with the raw data.

The analysis identified several prominent themes related to teachers' perceptions of AI-based feedback. These themes were categorized into two main groups: themes related to teachers' perceptions and themes related to the factors influencing their use of AI in feedback. The themes related to teachers' perceptions include views on the effectiveness of AI feedback, its impact on teaching

Table 2
Results of independent samples t-tests with welch correction for the experimental and control groups' pretest.

	t-test for Equality of means						
	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95 % Confidence Interval of the Difference	
						Lower	Upper
TR.pr	.32	65.89	.75	.05	.18	−0.30	.42
CC.pr	1.06	65.92	.29	.17	.16	−0.15	.50
LR.pr	.24	65.98	.80	.02	.11	−0.20	.26
GA.pr	.41	61.59	.68	.05	.14	−0.22	.34
OS.pr	1.06	64.47	.29	.35	.33	−0.30	1.01

Table 3
Descriptive statistics of the experimental and control groups for the TR, CC, LR, GA, and OS posttest results.

	Group	N	Mean	Std. Deviation	Std. Error Mean	Skewness Ratio	Kurtosis Ratio
TR.post	Experimental Group	34	3.67	.63	.10	−1.45	0.08
	Control Group	34	2.85	.60	.10	1.21	0.45
CC.post	Experimental Group	34	3.67	.72	.12	−1.01	- 1.51
	Control Group	34	2.44	.56	.09	1.69	1.33
LR.post	Experimental Group	34	3.29	.79	.13	0.09	0.77
	Control Group	34	2.47	.50	.08	0.99	0.06
GA.post	Experimental Group	34	3.41	.60	.10	0.25	1.39
	Control Group	34	2.64	.54	.09	−1.78	0.81
OS.post	Experimental Group	34	14.14	1.25	.21	1.91	−1.29
	Control Group	34	10.41	1.23	.21	1.33	0.04

Table 4
Results of independent samples t-test with welch correction for the experimental and control groups' posttest.

	t-test for Equality of Means						
	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95 % Confidence Interval of the Difference	
						Lower	Upper
TR.post	5.44	65.86	.00	.82	.15	.52	1.12
CC.post	7.84	62.00	.00	1.23	.15	.92	1.55
LR.post	5.07	55.85	.00	.82	.16	.49	1.14
GA.post	5.46	65.17	.00	.76	.14	.48	1.04
OS.post	12.35	65.97	.00	3.73	.30	3.13	4.33

Table 5
Results of paired samples t-test for the experimental group's pretest and posttest.

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95 % Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	TR.pr - TR.post	−1.26	.75	.12	−1.52	−1.00	−9.81	33	.00
Pair 2	CC.pr - CC.post	−1.67	.87	.15	−1.98	−1.37	−11.13	33	.00
Pair 3	LR.pr - LR.post	−0.91	.57	.09	−1.11	−0.71	−9.32	33	.00
Pair 1	GA.pr - GA.post	−0.85	.55	.09	−1.04	−0.65	−8.91	33	.00
Pair 2	OS.pr - OS.post	−4.76	1.63	.28	−5.33	−4.19	−17.00	33	.00

a. Group = Experimental Group.

practices, and concerns about potential drawbacks. Factors influencing their use of AI include training, digital literacy, and institutional policies. To preserve confidentiality, pseudonyms (Teacher1 [T1] to Teacher8 [T8]) were used to present excerpts from the interview data.

4.3. Benefits of AI writing feedback

The teachers mostly praised AI systems for their individualized and responsive characteristics. T1 described how efficient revision is facilitated by customizing feedback prompts to each student's specific problems, stating, "We tailored prompts for each learner in accordance with their particular weaknesses in academic writing and IELTS proficiency". T2 emphasized promptness, stating, "Real-time feedback enabled students to improve their work efficiently." This was something that numerous instructors perceived as a supplement to teacher feedback and a means to free up their time. Further, T3 elaborated on the adaptive capabilities, stating, "By identifying patterns in the writing errors of each student, the AI feedback is customized to their specific requirements."

4.4. Limitations of AI writing feedback

Significant concerns emerged, however, regarding overreliance, precision, the possibility of plagiarism, and a lack of nuance. The analysis was criticized by T4, who stated, "The AI feedback failed to account for significant contextual implications." Regarding the dangers of plagiarism associated with an excessive reliance on AI by students, T5 cautioned, "Students who rely heavily on ChatGPT must be monitored for plagiarism." T1 elaborated on the difficulty of guaranteeing precision, stating, "It is difficult to verify that all of its assessments and guidance are accurate." T5 further elaborated on the challenge of finding appropriate use, stating, "Educators necessitate targeted direction…to determine the suitable applications and constraints of AI writing feedback."

4.5. Implementation recommendations

As recommendations for addressing constraints, emphasis was placed on oversight, objective alignment, and the development of use policies. Teacher oversight is crucial, according to T6, so "AI must be aligned with learning objectives by providing context and parameters." T4 advocated for ethical usage training, asserting, "Professional development on…responsible application of AI is required for staff." T6 advocated for establishing academic integrity standards about these tools, stating, "Faculty should collaborate to update academic integrity codes in order to address appropriate AI use."

4.6. Teacher digital literacy perspectives

Instructors articulated the necessity to cultivate comprehension regarding AI writing tools' merits, demerits, and conscientious implementations. Professional development workshops that emphasize practical experience with educational AI tools are required, according to T7. "Teachers require focused guidance" regarding proper usage, as stated by T1 about boundaries. According to T7, to remain informed about the latest developments in AI instruction and research, educators must be provided with continuous learning opportunities.

4.7. Changing roles of teachers

"In order for students to develop independent skills and use AI responsibly, teachers must maintain regular involvement in the classroom," T8 explained regarding the continued necessity for participation. T1 elaborated on using experience to augment AI, stating, "Our specialized expertise empowers us to provide AI input with valuable, individualized feedback." T2 espoused the cause of accountability, saying, "Educators must proactively detect and alleviate the potential for plagiarism that may result from the unrestrained utilization of artificial intelligence."

4.8. Academic integrity concerns

T8 argued for revised policies, stating, "Faculty should work together to revise academic integrity codes to address appropriate AI usage." T8 stated, "Teachers significantly contribute to AI ethical norms via their instructional policies and standards to mitigate risks." "It is the responsibility of the educators to proactively identify and mitigate the risk of plagiarism caused by unrestrained AI use," stated T2, which identified reactive vigilance.

4.9. Impact on critical thinking

T5 emphasized the importance of pushing boundaries, stating, "Our responsibility is to push students beyond the capabilities of AI in order to improve their capacity for critical thinking." "Teachers must identify gaps in AI feedback and provide students with more meaningful and effective feedback," as stated in T3. T7 encouraged originality, stating, "Educators enable pupils to develop into innovative thinkers through an emphasis on originality in writing."

5. Discussion

The first research question examined whether integrating adaptive ChatGPT feedback with teachers' feedback leads to significant improvements in students' English writing skills. The results from the independent samples *t*-tests showed that there were no significant differences between the experimental and control groups in terms of task response (TR), cohesion and coherence (CC), lexical resource (LR), grammatical accuracy (GA), and overall score (OS) before the treatment, reflecting the initial challenges English language learners face in writing proficiency (Hasani, 2016). However, as indicated in Table 2, all the significance indices (two-tailed) were below 0.001 after the treatment, confirming that the experimental group significantly outperformed the control group in these areas. This suggests that integrating ChatGPT feedback with teacher feedback positively impacted students' writing skills, consistent with previous studies highlighting the potential of AI-powered feedback systems (Su et al., 2023). The paired samples t-Tests revealed that the experimental group showed substantial improvement in writing between the pretest and posttest, whereas the control group did not exhibit similar gains. These results support the hypothesis that ChatGPT, when integrated with teacher feedback, provides personalized writing support that significantly improves writing proficiency (Huang et al., 2023). Therefore, these findings emphasize the potential of AI-enhanced feedback systems, like ChatGPT, in fostering personalized learning experiences that can significantly improve students' writing skills.

The findings of this study are consistent with previous research examining the effects of peer and teacher feedback on ESL and EFL students' revision quality and writing performance. Erlam et al. (2013) demonstrated that feedback, especially teacher feedback, plays a crucial role in developing students' writing skills by enhancing their confidence and ability to revise effectively. These studies reinforce the importance of feedback in writing instruction and highlight its role in fostering writing development.

Additionally, the findings align with research comparing the effectiveness of teacher feedback and ChatGPT's AI-powered feedback (Dai et al., 2023; Hellas et al., 2023; Su et al., 2023). These studies have found that ChatGPT can provide relevant, detailed, and coherent feedback for tasks in areas such as data science, programming, and argumentative writing, highlighting the potential of AI in educational feedback (Hong, 2023; Kim et al., 2023; Liu & Ma, 2023). This study investigates the sustained effects of combining

AI-powered feedback with teacher feedback by examining the longer-term impact (i.e., into the following semester) of integrating ChatGPT with teacher feedback (Huang et al., 2023). This combined approach, or multi-approach, which integrates teacher feedback with AI feedback, provides strong, personalized feedback that effectively meets the individual needs of each student. Moreover, this study contributes uniquely to the literature by exploring teachers' perceptions and evaluations of ChatGPT feedback, an area that has received limited attention. This exploration sheds light on the critical role of educators' perspectives in shaping the future of AI integration in education (Al-Garaady & Mahyoob, 2023). Therefore, understanding teachers' viewpoints on AI-based feedback is crucial for maximizing its impact in educational settings and guiding future integration strategies.

In light of Sociocultural Theory (SCT), the findings of this study can be understood as an illustration of the mediated learning processes that enhance students' cognitive development within their Zone of Proximal Development (ZPD). Integrating AI-powered feedback from ChatGPT with traditional teacher feedback exemplifies how technological tools act as mediators to scaffold learning (Marginson & Dang, 2017). Through personalized and real-time feedback, ChatGPT supports students' writing skill development, enabling them to achieve higher proficiency levels than they could independently (Xu, 2024). This aligns with Vygotsky's emphasis on the importance of cultural tools in facilitating learning and cognitive growth (Marginson & Dang, 2017). However, the role of the teacher remains critical in contextualizing and augmenting AI feedback, addressing limitations such as the lack of nuance and contextual understanding inherent in AI systems (Rahimi, 2013). Teachers ensure that the integration fosters meaningful learning and critical thinking by providing complementary feedback and guiding responsible AI use. Thus, the combination of ChatGPT's technological affordances and teacher expertise exemplifies a sociocultural approach to creating an adaptive and collaborative learning environment that enhances students' writing skills while addressing the dynamic needs of diverse learners.

The second research question explored how teachers evaluated AI-assisted feedback for enhancing EFL writing skills. Thematic analysis of the interview data revealed several key perceptions and insights from teachers regarding the integration of ChatGPT with their feedback. Teachers recognized the benefits of AI feedback, mainly its customization, promptness, and adaptability. These benefits align with the idea that AI can support individualized learning, a principle emphasized by Lee et al. (2022), which aligns with the teaching needs of students with varying writing levels. AI's ability to provide instant feedback allows students to revise and improve their work more efficiently, contributing to a more engaging and responsive learning experience.

However, teachers also identified limitations and concerns regarding AI feedback. These included the AI's lack of contextual understanding, the potential for students to over-rely on the technology, and the risk of plagiarism. These concerns highlight the importance of teachers' roles in guiding students toward responsible use of AI in academic settings. Teachers expressed a strong need for professional development to understand AI tools and their limitations better. This aligns with the findings of Marzuki, Widiati, Rusdin, Darwin and Indrawati (2023), who suggest that educators need continuous digital literacy training to navigate the complexities of AI integration. Additionally, teachers emphasized that AI should complement, not replace, human feedback. This echoes the views of Guo & Wang, 2023, who highlight the importance of educators in ensuring the quality, accuracy, and ethical implications of feedback. Therefore, investing in teacher training and fostering a balanced integration of AI and human feedback are essential for maximizing the potential of AI-driven educational tools.

Teachers further noted that AI feedback, while valuable, must be critically engaged with to foster students' critical thinking and creativity. As T5 pointed out, educators need to push students beyond AI's capabilities to promote independent thinking development. This concern is particularly relevant in argumentative writing, where students must present original arguments and ideas. Teachers in this study emphasized that AI should not be seen as a replacement for critical thinking but rather as a tool that, when used appropriately, can enhance learning. This is consistent with Al-Garaady & Mahyoob, 2023, who argue that AI should serve as an enhancement rather than a substitute for human intellectual engagement. Therefore, fostering critical engagement with AI tools is essential to ensure students benefit from the technology and develop the creative and analytical skills required for academic success.

In addressing why teachers hold these perceptions, the thematic analysis suggests that their experiences with AI feedback led them to value the benefits of personalized and timely responses and made them aware of the challenges. These teachers' perceptions are informed by their deep understanding of writing pedagogy and their belief in the necessity of human oversight in AI-assisted learning environments. For example, teachers noted that while AI can provide detailed and consistent feedback, its lack of contextual sensitivity and ability to assess complex aspects of writing, such as nuance and voice, limits its effectiveness. Thus, teachers see themselves as integral to maintaining the educational integrity of AI feedback by supplementing it with their expertise and judgment.

Integrating ChatGPT in argumentative writing may pose some challenges and limitations. One of the main issues is authorship and plagiarism, as ChatGPT can produce original texts that may not reflect the students' ideas or words. This raises the need to reconsider the conventional definition of plagiarism, which may not be sufficient to address the use of AI-based writing tools (Dale & Viethen, 2021). Another concern is the quality and suitability of ChatGPT's feedback for argumentative writing. Since ChatGPT relies on statistical patterns learned from web data, its feedback may not always align with academic standards or meet the expectations for argumentative writing (Grassini, 2023; Wang & Demszky, 2023). Therefore, students and teachers should remain cautious and critical when using ChatGPT's feedback in argumentative writing, ensuring that it complements, rather than replaces, the nuanced feedback from human educators.

5.1. Limitations

There are, however, limitations to this study. The research was conducted over a relatively short 12 sessions, with only two sessions per week dedicated to argumentative essay instruction. This restricted the depth of the writing instruction and feedback students could receive. Additionally, the relatively small sample size may limit the generalizability of the findings.

5.2. Pedagogical implications

This study has significant implications for educators exploring or already integrating AI tools into their writing instruction and assessment. First, this study demonstrates that combining ChatGPT with teachers' feedback based on IELTS criteria can significantly improve students' writing skills, particularly in argumentative essays. This combination aligns with the recognized need for practical, adaptive feedback in writing instruction (Cao & Zhong, 2023). Teachers should, therefore, consider using ChatGPT as a supplementary tool that provides personalized, prompt, and adaptive feedback on students' writing. This combination of AI and human feedback can enhance students' writing skills and scores in specific writing tasks, offering a more assertive and more personalized approach than traditional feedback methods.

The thematic analysis of teachers' perceptions also highlights the importance of educators possessing digital literacy skills and ethical awareness when using ChatGPT. Teachers need professional development and training programs that focus on how to use ChatGPT responsibly, interpret AI-generated feedback, and integrate these tools seamlessly into their curriculum. These programs should ensure that educators are well-equipped to leverage AI tools like ChatGPT while maintaining their essential role in the classroom.

5.3. Suggestions for future research

Future research should consider extending the duration and frequency of workshops to encompass a broader range of IELTS writing tasks and essay types. This would allow for a more comprehensive exploration of how integrating ChatGPT with teacher feedback can enhance EFL students' writing skills. By diversifying instructional content and increasing engagement time, researchers can gain deeper insights into the efficacy of this combined approach in fostering writing proficiency.

6. Conclusion

In conclusion, this study reinforces the potential of integrating AI-powered feedback with traditional teacher feedback in enhancing students' writing development. The combination of ChatGPT's AI feedback, which provides personalized, adaptive support, and teachers' detailed, contextualized feedback tailored to individual student needs, offers a robust and constructive approach to writing instruction. This multi-approach feedback model improves writing skills and helps meet diverse student needs more effectively than relying on either feedback method alone. As AI tools continue to evolve, it is crucial educators must balance innovation with their expert understanding of student needs, ensuring that AI serves as a complement to, rather than a replacement for, human instruction. Further research should continue to explore how this integrated approach can be optimized for diverse educational contexts and writing tasks.

Ethics approval and consent to participate

The authors certify that this work has not been published, nor is it under consideration for publication in any other journal. The authors take responsibility for its content and for contributing to the conception, design, and execution of the work, analysis and interpretation of data, and participating in the drafting of the text and its revisions, as well as approving the final version submitted.

Consent for publication

Not applicable.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for profit sectors.

Authors' contributions

MA: As the principal author, MA played a pivotal role in the study. MA was responsible for the comprehensive design and methodology, meticulously conducting all procedures, and ensuring the integrity of the data collection. Additionally, MA authored the initial draft of the manuscript, laying the foundation for the study's narrative and analysis.

S E: S E, the supervising professor, provided essential oversight and guidance throughout the research process. S E critically reviewed the final draft, offering substantial revisions and insightful comments that significantly enhanced the manuscript's quality and rigor.

L M: L M made valuable contributions by collecting supplementary data and meticulously revising specific sections of the manuscript. L M's efforts ensured the accuracy and coherence of the final document.

CRediT authorship contribution statement

Marjan Asadi: Writing – review & editing, Writing – original draft, Validation, Supervision, Software, Resources, Project

administration, Methodology, Formal analysis, Data curation, Conceptualization. **Saman Ebadi:** Writing – review & editing, Supervision. **Laleh Mohammadi:** Writing – review & editing, Resources, Data curation.

Declaration of competing interest

The authors declare that they have no competing interests

Acknowledgements

No applicable.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.tsc.2025.101766.

Data availability

Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

References

Al-Garaady, J., & Mahyoob, M. (2023). ChatGPT's capabilities in spotting and analyzing writing errors experienced by EFL learners. Arab World English Journals. *Special Issue on CALL, 9*, 3–17.

Alkaissi, H., & McFarlane, S. I. (2023). Artificial hallucinations in ChatGPT: implications in scientific writing. *Cureus, 15*(2).

Asadi, M., & Taheri, R. (2024). Enhancing peer assessment and engagement in online IELTS writing course through a teacher's multifaceted approach and AI integration. *Technology Assisted Language Education, 2*(2), 94–117. https://doi.org/10.22126/tale.2024.11083.1058

Baidoo-Anu, D., & Ansah, L. O. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. *Journal of AI, 7*(1), 52–62.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101.

Cao, S., & Zhong, L. (2023). Exploring the effectiveness of ChatGPT-based feedback compared with teacher feedback and self-feedback: Evidence from Chinese to English translation. *arXiv preprint arXiv:2309.01645*.

Clarizia, F., Colace, F., Lombardi, M., Pascale, F., & Santaniello, D. (2018). Chatbot: An education support system for student. Paper presented at the. In *Cyberspace Safety and Security: 10th International Symposium, CSS 2018, Amalfi, Italy, October 29–31, 2018*. Proceedings 10.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Creswell, J. W., Clark, V. L. P., Gutmann, M. L., & Hanson, W. E. (2003). Advanced mixed. *Handbook of mixed methods in social & behavioral research* (pp. 209–240).

Cullen, P. (2020). The Key to IELTS writing Task 2 course: Bands 7, 8, 9.

Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications.

Cunningham, J. M. (2019). Composition students' opinions of and attention to instructor feedback. *Journal of Response to Writing, 5*(1), 3–22.

Dale, R., & Viethen, J. (2021). The automated writing assistance landscape in 2021. *Natural Language Engineering, 27*(4), 511–518.

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

Cunningham, K. J., & Link, S. (2021). Video and text feedback on ESL writing: Understanding attitude and negotiating relationships. *Journal of Second Language Writing, 52*, 100797. https://doi.org/10.1016/j.jslw.2021.100797

Dai, W., Lin, J., Jin, H., Li, T., Tsai, Y.-S., Gašević, D., et al. (2023). Can large language models provide feedback to students? A case study on ChatGPT. In *Paper presented at the 2023 IEEE International Conference on Advanced Learning Technologies (ICALT)*.

Dogan, M. E., Goru Dogan, T., & Bozkurt, A. (2023). The use of artificial intelligence (AI) in online learning and distance education processes: A systematic review of empirical studies. *Applied Sciences, 13*(5), Article 3056.

Erlam, R., Ellis, R., & Batstone, R. (2013). Oral corrective feedback on L2 writing: Two approaches compared. *System, 41*(2), 257–268.

Ferris, D. (2011). *Treatment of error in second language student writing*. University of Michigan Press.

Firat, M. (2023). How chat GPT can transform autodidactic experiences and open education. *International Journal of Educational Technology in Higher Education, 20*(1), 1–16. https://doi.org/10.31219/osf.io/9 ge8m

Fu, Q.-K., Zou, D., Xie, H., & Cheng, G. (2022). A review of AWE feedback: Types, learning outcomes, and implications. *Computer Assisted Language Learning, 37*(1–2), 179–221.

Gao, J. (2021). Exploring the feedback quality of an automated writing evaluation system pigai. *International Journal of Emerging Technologies in Learning (iJET), 16*(11), 322–330.

Gauvain, M. (2020). *Vygotsky's sociocultural theory. Encyclopedia of Infant and Early Childhood Development* (pp. 446–454). https://doi.org/10.1016/B978-012809324-5.23569-4

Godwin-Jones, R. (2019). In a world of SMART technology, why learn another language? *Journal of Educational Technology & Society, 22*(2), 4–13.

Grassini, S. (2023). Shaping the future of education: Exploring the potential and consequences of AI and ChatGPT in educational settings. *Education Sciences, 13*(7), Article 692.

Guo, K., & Wang, D. (2023). To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing. *Education and Information Technologies, 29*(7), 8435–8463.

Hasani, A. (2016). Enhancing argumentative writing skill through contextual teaching and learning. *Educational Research and Reviews, 11*(16), 1573–1578.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112.

Hellas, A., Leinonen, J., Sarsa, S., Koutcheme, C., Kujanpää, L., & Sorva, J. (2023). Exploring the responses of large language models to beginner programmers' help requests. *arXiv preprint arXiv:2306.05715*.

Herrington, A., & Moran, C. (2012). Writing to a machine is not writing at all. *Writing assessment in the 21st century: Essays in honor of edward M. white* (pp. 219–232). Cresskill: Hampton Press.

Hong, W. C. H. (2023). The impact of ChatGPT on foreign language teaching and learning: Opportunities in education and research. *Journal of Educational Technology and Innovation, 5*(1), 1–15.
Huang, J., & Tan, M. (2023). The role of ChatGPT in scientific communication: Writing better scientific review articles. *American Journal of Cancer Research, 13*(4), Article 1148.
Hyland, K., & Hyland, F. (2006). *Feedback on second language students' writing. Language teaching, 39*(2), 83–101.
Javaid, M., Haleem, A., & Singh, R. P. (2023). A study on ChatGPT for Industry 4.0: Background, potentials, challenges, and eventualities. *Journal of Economy and Technology, 1*, 127–143.
Karataş, F., & Yüce, E. (2024). AI and the Future of Teaching: Preservice teachers' reflections on the use of artificial intelligence in open and distributed learning. *The International Review of Research in Open and Distributed Learning, 25*(3), 304–325. https://doi.org/10.19173/irrodl.v25i3.7785
Kim, S., Shim, J., & Shim, J. (2023). A Study on the utilization of OpenAI ChatGPT as a second language learning tool. *Journal of Multimedia Information System, 10*(1), 79–88.
Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal, 54*(2), 537–550.
Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for *t*-tests and ANOVAs. *Frontiers in Psychology, 4*, Article 863.
Lang, S., Aull, L., & Marcellino, W. (2019). A taxonomy for writing analytics. *The Journal of Writing Analytics, 3*, 13–37.
Lee, M., Liang, P., & Yang, Q. (2022). Coauthor: Designing a human-ai collaborative writing dataset for exploring language model capabilities. Paper presented at the. In *Proceedings of the 2022 CHI conference on human factors in computing systems*.
Li, C., Wei, L., & Lu, X. (2023). Contributions of foreign language writing emotions to writing achievement. *System, 116*, Article 103074.
Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic Inquiry* (pp. 1–416). Sage.
Liu, G., & Ma, C. (2023). Measuring EFL learners' use of ChatGPT in informal digital learning of English based on the technology acceptance model. *Innovation in Language Learning and Teaching, 18*(2), 125–138.
Liu, Q., & Wu, S. (2019). Same goal, varying beliefs: How students and teachers see the effectiveness of feedback on second language writing. *Journal of Writing Research, 11*(2), 299–330.
Liu, Y., Han, T., Ma, S., Zhang, J., Yang, Y., Tian, J., et al. (2023). Summary of chatgpt/gpt-4 research and perspective towards the future of large language models. *Meta-Radiology, 1*(2), Article 100017. https://doi.org/10.1016/j.metrad.2023.100017
Lu, Q., Qiu, B., Ding, L., Xie, L., & Tao, D. (2023). Error analysis prompting enables human-like translation evaluation in large language models: A case study on chatgpt. *arXiv preprint arXiv:2303.13809*.
Marginson, S., & Dang, T. (2017). Vygotsky's sociocultural theory in the context of globalization. *Asia Pacific Journal of Education, 37*(1), 116–129. https://doi.org/10.1080/02188791.2016.1216827
Marzuki, Widiati, U., Rusdin, D., Darwin, & Indrawati, I. (2023). The impact of AI writing tools on the content and organization of students' writing: EFL teachers' perspective. *Cogent Education, 10*(2), Article 2236469.
Mohammadi, M., Zarrabi, M., & Kamali, J. (2023). Formative assessment feedback to enhance the writing performance of Iranian IELTS candidates: Blending teacher and automated writing evaluation. *International Journal of Language Testing, 13*(1), 206–224.
Pallant, J. (2020). SPSS survival manual (7th ed.). Open University Press.Pallant, J. (2020). *SPSS survival manual: A step by step guide to data analysis using IBM SPSS*. Routledge.
Panofsky, C. P. (2003). The Relations of Learning and Student Social Class: Toward Re-"socializing" Sociocultural Learning Theory. In A. Kozulin, B. Gindis, V. S. Ageyev, & S. M. Miller (Eds.), *Vygotsky's Educational Theory in Cultural Context* (pp. 411–431). Cambridge: Cambridge University Press.
Potter, A., & Wilson, J. (2021). Statewide implementation of automated writing evaluation: Analyzing usage and associations with state test performance in grades 4-11. *Educational Technology Research and Development, 69*, 1557–1578.
Rahimi, M. (2013). Is training student reviewers worth its while? A study of how training influences the quality of students' feedback and writing. *Language Teaching Research, 17*(1), 67–89. https://doi.org/10.1177/1362168812459151
Roscoe, R. D., Allen, L. K., Johnson, A. C., & McNamara, D. S. (2018). Automated writing instruction and feedback: Instructional mode, attitudes, and revising. Paper presented at the. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*.
Shabani, K. (2016). Applications of Vygotsky's sociocultural approach for teachers' professional development. *Cogent Education, 3*(1), 1252177.
Su, Y., Lin, Y., & Lai, C. (2023). Collaborating with ChatGPT in argumentative writing classrooms. *Assessing Writing, 57*, Article 100752.
Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics*. Boston, MA: Pearson Education.
Taskiran, A., & Goksel, N. (2022). Automated feedback and teacher feedback: Writing achievement in learning English as a foreign language at a distance. *Turkish Online Journal of Distance Education, 23*(2), 120–139.
Waer, H. (2023). The effect of integrating automated writing evaluation on EFL writing apprehension and grammatical knowledge. *Innovation in Language Learning and Teaching, 17*(1), 47–71.
Wei, P., Wang, X., & Dong, H. (2023). The impact of automated writing evaluation on second language writing skills of Chinese EFL learners: A randomized controlled trial. *Frontiers in Psychology, 14*(2), 345–357.
Wang, R.E., .& Demszky, D. (2023). Is ChatGPT a Good Teacher Coach? Measuring Zero-Shot Performance For Scoring and Providing Actionable Insights on Classroom Instruction. *arXiv preprint* arXiv:2306.03090.
Wilson, J., & Czik, A. (2016). Automated essay evaluation software in English Language Arts classrooms: Effects on teacher feedback, student motivation, and writing quality. *Computers & Education, 100*, 94–109.
Xu, W., Meng, J., Raja, S. K. S., Priya, M. P., & Kiruthiga Devi, M. (2023). Artificial intelligence in constructing personalized and accurate feedback systems for students. *International Journal of Modeling, Simulation, and Scientific Computing, 14*(01), Article 2341001.
Yüce, E. (2019). Possible problems in online foreign language teaching at a university context. *International Journal of Curriculum and Instruction, 11*(2), 75–86.
Yüce, E. (2022). The immediate reactions of EFL learners towards total digitalization at higher education during the Covid-19 pandemic. *Journal of Theoretical Educational Science, 15*(1), 1–15. https://doi.org/10.30831/akukeg.939836
Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education–where are the educators? *International Journal of Educational Technology in Higher Education, 16*(1), 1–27.
Zhang, J., & Zhang, L. J. (2024). The effect of feedback on metacognitive strategy use in EFL writing. *Computer Assisted Language Learning, 37*(5–6), 1198–1223.
Zhang, Z. V., & Hyland, K. (2018). Student engagement with teacher and automated feedback on L2 writing. *Assessing Writing, 36*, 90–102.
Zhu, M., Liu, O. L., & Lee, H.-S. (2020). The effect of automated feedback on revision behavior and learning gains in formative assessment of scientific argument writing. *Computers & Education, 143*, Article 103668.
Xu, S. (2024). An In-depth examination of the sociocultural framework-informed practical model of scaffolded feedback in enhancing senior high school english writing instruction. *International Journal of Social Sciences and Public Administration, 3*(3), 110–119. https://doi.org/10.62051/ijsspa.v3n3.13