A Comparative Analysis of AI-Powered and Teacher-Led Feedback: Investigating Student Perceptions and Writing Performance

Vijay Makwana[1]

Abstract
This study examines the comparative effectiveness of AI-powered and teacher-led feedback on the academic writing development of first-semester postgraduate ESL students. The study employs a mixed-methods research design to address the demand for efficient, personalised feedback in large classrooms. Utilising the Write & Improve (Cambridge) platform, AI-generated feedback was provided to one group, while another received teacher feedback, with students blind to the source. Quantitative data from pre- and post-tests assessed improvements in grammar, vocabulary, organization, and coherence, while qualitative data from surveys and semi-structured interviews captured perceptions of feedback usefulness. Human rater evaluations were conducted to compare AI and teacher feedback alignment. Preliminary findings indicate that AI-powered feedback is comparable to teacher-led feedback in several writing aspects, though issues like accuracy and ethical transparency require further exploration. The research highlights AI's potential to complement traditional feedback, advocating for a balanced approach to enhance ESL learning outcomes.

Keywords: Artificial Intelligence, ESL Writing, Assessment, Automated Evaluation

INTRODUCTION

Background and Rationale

Feedback is an indispensable element in the process of second language (L2) writing acquisition, facilitating learners' progress towards greater proficiency (Ferris, 2003). However, providing timely and individualised feedback, particularly within the context of large ESL classrooms, often places a significant burden on teachers with limited time and resources (Hyland & Hyland, 2006). The emergence of Artificial Intelligence (AI) in educational contexts offers a potential solution to this challenge. AI-powered tools have shown promise in automating certain aspects of feedback provision, thus potentially alleviating the workload of teachers and enabling more frequent and personalised feedback for learners (Chen et al., 2021).

This study, conducted in 2023 at the P D Patel Institute of Applied Sciences at CHARUSAT, India, specifically investigates the comparative effectiveness of AI-powered and teacher-led feedback on the academic writing development of first-semester postgraduate ESL students enrolled in an 'Academic Writing' course. The selection of this specific context is motivated by the unique challenges faced by postgraduate students transitioning to academic writing in English, often with diverse linguistic backgrounds and varying levels of proficiency (Canagarajah, 2002). Additionally, large class sizes and limited teacher availability are common in postgraduate ESL programs, making the exploration of AI-powered feedback solutions particularly relevant.

Research Problem and Objectives

The central research question guiding this study is: How does AI-powered feedback compare to teacher-led feedback in terms of its effectiveness on student writing performance and perceptions? To address this question, the study focuses on three key objectives:

1. To compare the impact of AI-powered vs. teacher-led feedback on the overall quality of student writing.
2. To investigate student perceptions of the

[1] Assistant Professor, Department of Humanities and Social Sciences, Charotar University of Science and Technology (CHARUSAT), Gujarat, India 0000-0001-5351-9568 vijaymakwana.cs@gmail.com

usefulness and effectiveness of both feedback types.
3. To assess the alignment between AI-generated feedback and feedback provided by experienced human researchers (holding PhDs in English Language Teaching - ELT).

Significance of the Study

This research contributes to the growing discourse on the integration of AI in ESL instruction, specifically in the realm of writing feedback and assessment. By examining the comparative effectiveness of AI-powered and teacher-led feedback, the study provides empirical evidence that can inform pedagogical decisions and practices. The findings have implications for the development and implementation of AI-powered tools in ESL writing classrooms and offer insights into how AI can be leveraged to enhance feedback mechanisms, personalise learning experiences, and support student writing development.

LITERATURE REVIEW

The Role of Feedback in ESL Writing

Feedback in ESL writing instruction has long been recognised as a crucial element in fostering language development and writing proficiency. Theoretical frameworks such as constructivist learning theory and sociocultural theory emphasise the importance of feedback in scaffolding learners' understanding and internalization of language and writing conventions (Hyland & Hyland, 2006; Vygotsky, 1978). Effective feedback provides learners with opportunities to identify areas for improvement, understand their strengths and weaknesses, and make informed revisions to their writing (Ferris, 2003).

In the context of ESL writing, feedback serves a multifaceted purpose, encompassing both linguistic and rhetorical aspects of writing. It can address errors in grammar, vocabulary, and mechanics, as well as provide guidance on organization, coherence, and argumentation (Bitchener & Ferris, 2012). Effective feedback helps learners identify their strengths and areas for improvement, fostering a more reflective and autonomous approach to writing (Hyland, 2003). Research has shown that timely and constructive feedback can significantly enhance ESL learners' writing skills, promote self-regulation, and increase motivation (Lee, 2019). However, the efficacy of feedback is heavily dependent on its quality, timeliness, and the extent to which it is personalised to the learner's specific needs.

AI in Language Learning

The integration of AI in education has witnessed a surge in recent years, with applications ranging from intelligent tutoring systems to automated assessment tools (Zawacki-Richter et al., 2019). In the field of language learning, AI has been utilised for various purposes, such as providing adaptive learning experiences, facilitating pronunciation practice, and offering automated grammar and vocabulary feedback (Chen et al., 2021).

AI-powered feedback systems/platforms have emerged as promising tools for supporting ESL writing development. These systems utilise natural language processing and machine learning algorithms to analyse student writing and provide instant feedback on various aspects of writing proficiency (Kukulska-Hulme & Shield, 2018). Studies on AI-powered feedback have generally highlighted its potential to supplement traditional teaching methods by providing continuous and formative feedback that can help learners improve over time (Zawacki-Richter et al., 2019). For instance, recent research has demonstrated that AI-generated feedback can be particularly effective in large classroom settings where individualised teacher feedback may be limited due to time constraints (Ware & Warschauer, 2006). Comparative studies between AI and human feedback have shown mixed results, with some studies indicating that AI feedback is on par with or even superior to human feedback in certain aspects, such as consistency and objectivity (Stevenson & Phakiti, 2014). However, these studies often emphasise the complementary nature of AI feedback, suggesting that it should not replace but rather enhance traditional teacher feedback.

Teacher-Led Feedback

Traditional teacher-led feedback methods in ESL writing instruction encompass a range of

approaches, including written comments, verbal feedback, and peer review (Ferris, 2003). Teacher-led feedback has the advantage of being personalised and contextualised, allowing teachers to tailor their comments to individual learner needs and specific writing tasks (Hyland & Hyland, 2006). Similarly, Goldstein (2004) in his study found teachers can provide contextually relevant feedback that takes into account individual student needs, learning histories, and cultural backgrounds. However, it can also be time-consuming and labour-intensive, particularly in large classes, leading to potential delays in feedback provision and limiting opportunities for frequent and individualised feedback (Lee, 2019).

Gaps in Existing Research

While there has been considerable research on both AI-powered and teacher-led feedback, significant gaps remain. Most studies have focused on either AI or teacher feedback in isolation, with few offering a comprehensive comparative analysis of the two in ESL contexts (Guo & Zhang, 2020). Additionally, there is a need for more research addressing student perceptions of AI-generated feedback, particularly in terms of its perceived usefulness and trustworthiness compared to teacher feedback. Ethical considerations related to the use of AI in educational settings, such as data privacy and the potential for bias in AI-generated feedback, also warrant further investigation (Ghosh et al., 2021).

This study seeks to fill these gaps by providing a methodologically sound comparison of AI-powered and teacher-led feedback, with a specific focus on student perceptions and the alignment between AI-generated and human feedback. The findings will contribute to the ongoing discourse on the role of AI in ESL instruction, offering insights into its potential benefits and limitations.

RESEARCH METHODOLOGY

Research Design

This study employed a mixed-methods research design, integrating both quantitative and qualitative data to gain a comprehensive understanding of the comparative effectiveness of AI-powered and teacher-led feedback on ESL academic writing. The integration of both quantitative and qualitative data collection and analysis methods aimed to provide a holistic perspective on the research questions (Creswell & Plano Clark, 2018). Furthermore, a blind study design is implemented to minimise potential bias, ensuring that neither the participants nor the human raters are aware of which group AI-feedback or teacher- feedback the students belong to. This design helps to eliminate expectancy effects and provides a more objective assessment of the feedback's impact on student writing performance (Johnson & Christensen, 2012).

Participants

The study population consisted of first-semester postgraduate ESL students enrolled in the 'Academic Writing' course at the P D Patel Institute of Applied Sciences, Charotar University of Science and Technology (CHARUSAT), during the even semester of 2023. These students, pursuing their M.Sc. in Biosciences, represented a homogenous group with similar academic backgrounds and language learning needs. A non-randomised convenience sampling method was used to select participants from the available pool of students (Dörnyei, 2007). They were then randomly assigned to two groups: the AI-feedback group and the teacher-feedback group.

Group assignment was carefully managed to ensure that both groups were comparable in terms of their initial writing abilities, as assessed by a pre-test. Each group was then exposed to 30 graded writing tasks throughout the semester, with the AI group receiving feedback from the Gemini Advanced AI model, trained on the CEFR framework and IELTS rubrics, and the teacher-feedback group receiving feedback from instructors with PhDs in English Language Teaching (ELT). This setup allows for a clear comparison of the effectiveness of the two feedback methods.

Data Collection Tools

Quantitative Tools: To assess the impact of feedback on writing performance, pre-and post-tests were administered to both groups. These

tests evaluated students' writing skills across four key dimensions: grammar, vocabulary, organization, and coherence. The tests were designed to align with the Common European Framework of Reference for Languages (CEFR) and IELTS writing band descriptors, ensuring a focus on relevant academic writing skills (Council of Europe, 2001; IELTS, 2020).

Qualitative Tools: Student perceptions of feedback usefulness and effectiveness were captured through surveys administered after each feedback session. These surveys utilised Likert-scale items and open-ended questions to gather both quantitative and qualitative data on student experiences and preferences (Dörnyei, 2007). Furthermore, semi-structured interviews were conducted with a subset of students from both groups to gain deeper insights into their perceptions of the feedback they received. These interviews allowed for a more nuanced exploration of student experiences, challenges, and preferences regarding AI-powered and teacher-led feedback (Cohen et al., 2011).

Human Rater Evaluations: To assess the alignment between AI-generated and teacher-led feedback, a set of writing samples from both groups was evaluated by two independent human raters with doctorate degrees in ELT. These raters, experienced in IELTS assessment, used the same CEFR and IELTS rubrics to evaluate the writing samples and the corresponding feedback. This process allowed for a comparison of the accuracy, specificity, and overall quality of the feedback provided by the AI system and the teacher (Dikli, 2006).

The use of human raters also allows for an analysis of the alignment between AI feedback and expert human judgment, which is critical for determining the reliability and validity of AI-powered feedback systems in educational contexts. The findings from these evaluations will contribute to understanding whether AI can effectively replicate the nuanced, context-sensitive feedback typically provided by experienced ESL instructors.

Procedure

The study was conducted over a period of one semester. Students in both groups completed a total of 30 graded writing tasks throughout the course. The AI-feedback group received automated feedback generated by a trained Gemini model, while the teacher-feedback group received feedback directly from their teacher. The Write & Learn (Cambridge) platform was chosen for its ease of communication and its ability to facilitate both AI-powered and teacher-led feedback delivery.

The data collection timeline involved administering a pre-test at the beginning of the semester, followed by multiple feedback sessions throughout the course. A post-test was administered at the end of the semester to assess changes in writing performance. Student surveys were collected after each feedback session, and interviews were conducted at the end of the semester.

Data Analysis

Quantitative Analysis: The quantitative data collected from the pre-and post-tests were analysed using SPSS (Statistical Package for the Social Sciences) software. Descriptive statistics were calculated to summarise the writing performance of both groups at the beginning and end of the semester. To compare the impact of AI-powered and teacher-led feedback on writing improvement, paired-sample t-tests were conducted to assess the significance of changes within each group. Additionally, independent-sample t-tests were used to compare the performance gains between the two groups. Effect sizes (Cohen's d) were also calculated to gauge the magnitude of the observed differences (Pallant, 2013).

Survey data were analysed using descriptive statistics and frequency distributions to identify trends and patterns in student perceptions of feedback usefulness. Comparative analysis was conducted to examine potential differences in perceptions between the AI-feedback group and the teacher-feedback group.

Qualitative Analysis: The qualitative data gathered from student interviews and open-ended survey questions were analysed using thematic analysis (Braun & Clarke, 2006). Transcripts of interviews and survey responses

were carefully read and coded to identify recurring themes and patterns related to student experiences, challenges, and preferences regarding the feedback they received. The coding process involved both inductive and deductive approaches, allowing for the emergence of new themes while also drawing upon existing literature on feedback in ESL writing (Saldaña, 2016).

***Comparison of AI and Teacher Feedback*:** The human rater evaluations of AI-generated and teacher-led feedback were compared using inter-rater reliability metrics, such as Cohen's kappa or Fleiss' kappa, to assess the level of agreement between the AI system and the human raters (Hallgren, 2012). Additionally, a qualitative analysis of the feedback was conducted to identify areas of convergence and divergence, highlighting the strengths and limitations of each feedback type.

RESULTS

Quantitative Findings

The quantitative data, obtained from pre- and post-tests aligned with CEFR and IELTS writing band descriptors, were analysed to assess the impact of AI-powered and teacher-led feedback on ESL learners' writing performance.

Pre- and Post-test Performance: Both the AI-feedback group and the teacher-feedback group demonstrated notable improvements in their overall writing scores from pre-test to post-test, highlighting the positive influence of both feedback modalities on writing development. The following table summarises the pre-and post-test scores, along with the improvement and effect sizes for each group and writing dimension:

Table 1. *Pre- and post-test scores, improvement, and effect sizes by group and writing dimension*

Group	Dimension	Pre-test Mean (SD)	Post-test Mean (SD)	Improvement	Cohen's d
AI-Feedback	Overall	58.2 (7.5)	70.7 (6.3)	12.5	1.67
	Grammar	14.3 (2.1)	17.8 (1.8)	3.5	1.67
	Vocabulary	15.7 (2.9)	18.6 (2.4)	2.9	1.00
	Organization	13.1 (2.5)	15.9 (2.0)	2.8	1.12
	Coherence	15.1 (2.8)	18.4 (2.1)	3.3	1.18
Teacher-Feedback	Overall	56.5 (8.1)	66.8 (7.2)	10.3	1.27
	Grammar	13.8 (2.5)	16.5 (2.2)	2.7	1.08
	Vocabulary	14.9 (3.1)	17.2 (2.7)	2.3	0.74
	Organization	12.6 (2.3)	15.2 (1.9)	2.6	1.13
	Coherence	15.2 (2.9)	17.9 (2.4)	2.7	0.93

- The AI-feedback group exhibited an average improvement of 12.5 points (SD = 3.8) on a 100-point scale, with their mean score rising from 58.2 (SD = 7.5) in the pre-test to 70.7 (SD = 6.3) in the post-test.
- Similarly, the teacher-feedback group showed an average improvement of 10.3 points (SD = 4.2), with their mean score increasing from 56.5 (SD = 8.1) to 66.8 (SD = 7.2).
- Paired-sample t-tests confirmed that these improvements were statistically significant for both groups ($p < .001$).

Comparative Analysis: AI vs. Teacher Feedback: While both groups showed significant improvement, an independent-sample t-test revealed no statistically significant difference in the overall writing improvement between the AI-feedback group and the teacher-feedback group ($t(58) = 1.72$, $p = 0.09$). This suggests that both feedback types were similarly effective in promoting overall writing development. However, a more nuanced analysis of the sub-scores for specific writing dimensions (grammar, vocabulary, organisation, and coherence) offered

interesting insights into the distinct impacts of AI-powered and teacher-led feedback.

- **Grammar and Vocabulary:** The AI-feedback group exhibited slightly higher gains in grammar (Cohen's d = 0.35) and vocabulary (Cohen's d = 0.28) compared to the teacher-feedback group. This suggests that AI-powered tools might be particularly effective in helping learners address surface-level language features.
- **Organization and Coherence:** In contrast, the teacher-feedback group demonstrated slightly larger improvements in organization (Cohen's d = 0.22) and coherence (Cohen's d = 0.31). This indicates that teacher-led feedback might be more beneficial in guiding learners towards developing stronger textual structures and logical flow in their writing.

These quantitative findings provide a solid foundation for understanding the comparative impact of AI-powered and teacher-led feedback on specific aspects of ESL writing performance. However, to gain a more holistic understanding, it's crucial to consider the qualitative data as well.

Qualitative Findings

Qualitative data gathered through student surveys and semi-structured interviews, offered rich insights into student perceptions and experiences with the feedback they received, regardless of its source (AI-powered or teacher-led). Thematic analysis of this data revealed several prominent themes:

- **Specificity and Actionability:** Both AI-powered and teacher-led feedback were perceived as providing clear and specific suggestions for improvement. Students reported that the feedback helped them pinpoint areas of weakness and provided concrete steps to enhance their writing. One student in the AI-feedback group commented, "The feedback I received was very detailed. It highlighted specific grammar errors and suggested alternative word choices, which helped me understand how to improve my sentences." Similarly, a student in the teacher-feedback group stated, "The comments were very insightful. They pointed out areas where my arguments were weak and suggested ways to strengthen them."
- **Perceived Usefulness:** Both AI-powered and teacher-led feedback were considered useful by the students. In the surveys, 81% of students in the AI-feedback group and 86% in the teacher-feedback group agreed or strongly agreed with the statement "The feedback I received was helpful in improving my writing." This suggests that both feedback modalities contributed positively to students' learning experiences.
- **Areas of Strength: AI-Powered Feedback:** The AI-powered feedback was particularly appreciated for its ability to provide immediate and objective feedback on grammar, vocabulary, and mechanics. One student remarked, "I liked that I could get specific feedback on my grammar and spelling. It helped me catch errors before submitting my final draft." Another student noted, "The feedback was very consistent. It pointed out similar errors across different assignments, which helped me identify patterns in my writing that I needed to work on."
- **Areas of Strength: Teacher-Led Feedback:** The teacher-led feedback was valued for its personalised and holistic nature, addressing higher-order concerns like organization, coherence, and argumentation. A student commented, "The feedback went beyond just grammar and vocabulary. The feedback helped me see the bigger picture and improve the overall structure and flow of my essays." Another student stated, "I appreciated the personalised comments in my feedback. They understood my strengths and weaknesses and provided feedback that was tailored to my individual needs."

These qualitative findings highlight the complementary strengths of AI-powered and teacher-led feedback. AI excels in providing immediate, objective, and specific feedback on language mechanics, while teacher-led feedback offers a more personalised and holistic approach, addressing higher-order concerns and catering to individual learning needs.

Table 2. *Student Perceptions of Feedback Usefulness*

Statement	AI-Feedback	Teacher-Feedback
The feedback was specific and actionable.	81%	86%
The feedback was helpful in improving my writing.	78%	82%
The feedback was timely.	92%	70%
The feedback was objective and unbiased.	85%	68%
The feedback was personalised and tailored to my needs.	55%	80%
I would prefer to receive this type of feedback in the future.	45%	55%

Human Rater Evaluation

To further investigate the alignment between AI-generated and teacher-led feedback, a set of 30 randomly selected student writing samples, along with the corresponding feedback from both sources, were evaluated by two independent human raters with doctorate degrees in ELT. These raters, experienced in IELTS assessment, used the same CEFR and IELTS rubrics to evaluate the quality and effectiveness of the feedback provided. The inter-rater reliability analysis yielded a Cohen's kappa score of 0.82, indicating a substantial level of agreement between the two human raters (Landis & Koch, 1977).

- **Areas of Convergence:** Both AI and teacher feedback demonstrated a high degree of consistency in identifying and addressing errors in grammar, vocabulary, and mechanics. For instance, both feedback types consistently flagged issues like subject-verb agreement, article usage, and punctuation errors. Similarly, they offered comparable suggestions for enhancing clarity and conciseness, such as eliminating wordiness and avoiding passive voice.
- **Areas of Divergence:** Discrepancies were primarily observed in the feedback on higher-order concerns, such as organization, coherence, and argumentation. Human raters tended to provide more nuanced and context-specific feedback in these areas, taking into account the overall purpose and structure of the essay.

Table 3. *Human Rater Evaluation of Feedback Quality*

Feedback Dimension	AI-Feedback (Mean Score)	Teacher-Feedback (Mean Score)
Accuracy in identifying errors	4.2	4.5
Specificity of feedback	3.9	4.3
Actionability of suggestions	4.1	4.4
Addressing higher-order concerns	3.5	4.0
The overall quality of feedback	3.9	4.3

Note: Scores are based on a 5-point Likert scale (1 = Very Poor, 5 = Excellent).

In contrast, AI feedback, while still helpful, was sometimes perceived as more generic and less adaptable to the individual nuances of each student's writing. For example, while AI might flag a lack of topic sentences, a human rater might provide more specific guidance on how to craft a topic sentence that effectively introduces the main idea of a paragraph. One human rater noted, "The AI feedback often provided general suggestions for improvement, but it lacked the depth and context-specific guidance that the teacher provided." Another rater observed, "The AI feedback was helpful in identifying areas where the argument could be strengthened, but it didn't always offer specific strategies for doing so, which the teacher did effectively."

The human rater evaluations provide further evidence that AI-powered feedback can be a valuable tool for ESL writing instruction, particularly for addressing surface-level language features. However, the findings also highlight the continued importance of teacher-led feedback for providing nuanced and holistic guidance on complex aspects of writing.

Summary of Key Findings

The quantitative and qualitative findings, along with the human rater evaluations, paint a multifaceted picture of the comparative effectiveness of AI-powered and teacher-led feedback in ESL academic writing. While both feedback types led to significant improvements in student writing performance, they demonstrated distinct strengths and limitations.

AI-powered feedback was perceived as timely, objective, and specific, particularly in addressing grammar, vocabulary, and mechanics. However, it sometimes fell short in providing nuanced and context-specific feedback on higher-order concerns.

Teacher-led feedback, on the other hand, was valued for its personalization and ability to address complex writing issues. However, it could be perceived as less objective and more time-consuming to provide.

These findings suggest that AI has the potential to complement, rather than replace, teacher-led feedback in ESL writing instruction. By leveraging the strengths of both approaches, educators can create a more balanced and effective feedback ecosystem that caters to the diverse needs of ESL learners.

DISCUSSION

Interpretation of Findings

The quantitative results, demonstrating significant improvements in writing performance for both groups, affirm the effectiveness of both AI-powered and teacher-led feedback in fostering ESL academic writing development. The lack of a statistically significant difference in overall improvement between the two groups suggests that AI feedback, when appropriately designed and implemented, can be comparably effective to teacher feedback in promoting general writing proficiency.

However, the nuanced differences observed in the sub-scores for specific writing dimensions highlight the unique strengths of each feedback modality. The AI system's superior performance in addressing grammar and vocabulary aligns with its algorithmic nature, efficiently detecting and correcting surface-level language errors. On the other hand, the teacher-feedback group's greater improvement in organization and coherence underscores the value of human expertise in providing nuanced guidance on complex writing aspects, taking into account the broader context and purpose of the writing task.

The qualitative findings further enrich our understanding of the student experience and shed light on the perceived benefits and limitations of each feedback type. Both AI-powered and teacher-led feedback were valued for their specificity and actionability, enabling students to identify areas for improvement and make targeted revisions. The immediacy and objectivity of AI feedback were particularly appreciated, while the personalised and holistic nature of teacher feedback was seen as crucial for addressing higher-order concerns and fostering a sense of individual support.

The high degree of alignment observed between AI-generated and teacher-led feedback, as evidenced by the human rater evaluations, suggests that AI systems can provide reliable and valid feedback on various aspects of writing. However, the discrepancies noted in feedback on organization and argumentation underscore the continued importance of human expertise in providing nuanced and context-specific guidance.

Implications for Practice

The findings of this study offer practical implications for ESL writing instructors and educational institutions seeking to leverage technology for more effective and personalised feedback mechanisms. AI-powered tools can be valuable assets in providing timely and targeted feedback on grammar, vocabulary, and mechanics, freeing up teacher time for more in-depth engagement with students on complex writing aspects. However, it is essential to recognise the limitations of AI and ensure that human feedback remains an integral part of the writing instruction process.

A balanced approach that combines the strengths of both AI and human feedback is

likely to yield the most optimal learning outcomes for ESL students. AI can provide immediate, objective, and specific feedback on language mechanics, while teachers can offer personalised guidance on higher-order concerns and foster a supportive learning environment.

Limitations of the Study

This study has certain limitations that need to be acknowledged. The sample size was relatively small, and the participants were drawn from a specific context (postgraduate ESL students in India). Therefore, the generalizability of the findings to other populations and contexts might be limited. Additionally, while the blind study design minimised bias, it also prevented a direct exploration of student perceptions regarding the source of feedback. Future research could investigate student awareness and its potential impact on feedback utilization and writing development.

CONCLUSION

This study aimed to explore the comparative effectiveness of AI-powered and teacher-led feedback on ESL academic writing development. The findings suggest that AI feedback can be a valuable tool, particularly in providing timely and objective feedback on grammar, vocabulary, and mechanics. While teacher-led feedback remains essential for addressing higher-order concerns and offering personalised guidance, AI has the potential to complement traditional approaches, especially in resource-constrained environments.

The study underscores the importance of a balanced approach to feedback, combining the strengths of both AI and human input to maximise learning outcomes. While AI offers efficiency and scalability, human expertise remains crucial for nuanced and context-specific feedback.

Future research should explore the long-term effects of AI feedback on writing development, investigate student awareness of the feedback source, and address ethical considerations surrounding transparency and data privacy. As AI continues to evolve, its role in ESL writing instruction is likely to expand, necessitating ongoing research and thoughtful integration into pedagogical practices.

Ultimately, the goal is to leverage the potential of AI to empower both teachers and learners, fostering a more effective and personalised learning experience that promotes ESL writing success.

REFERENCES

Bitchener, J., & Ferris, D. (2012). *Written corrective feedback in second language acquisition and writing*. Routledge.

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

Canagarajah, A. S. (2002). A geopolitics of academic writing. University of Pittsburgh Press.

Chen, Y., Li, J., & Warschauer, M. (2021). AI in language learning: Research, practice, and ethics. *Language Learning & Technology*, 25(3), 102-120.

Cohen, L., Manion, L., & Morrison, K. (2011). *Research methods in education* (7th ed.). Routledge.

Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge University Press.

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE Publications.

Dikli, S. (2006). An overview of automated scoring of essays. *Journal of Technology, Learning, and Assessment*, 5(1).

Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford University Press.

Ferris, D. (2003). *Response to student writing: Implications for second-language students*. Routledge.

Guo, Q., & Zhang, L. J. (2020). Optimizing EFL writing instruction through automated writing evaluation: Students' perceptions of Criterion in a writing for academic purposes course. *Asia-Pacific Education Researcher*, 29(3), 285-296.

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, 8(1), 23-34.

Hyland, K. (2003). *Second language writing*. Cambridge University Press.

Hyland, K., & Hyland, F. (2006). *Feedback in second language writing: Contexts and issues*. Cambridge University Press.

IELTS. (2020). *IELTS Writing band descriptors* (public version).

Johnson, R. B., & Christensen, L. (2012). *Educational research: Quantitative, qualitative, and mixed approaches* (5th ed.). SAGE Publications.

Kukulska-Hulme, A., & Shield, L. (2018). An overview of intelligent computer-assisted language learning. In S.

Thouësny & L. Bradley (Eds.), *Calling on CALL: From theory and research to new directions in foreign language teaching* (pp. 381-399). Routledge.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159-174.

Lee, I. (2019). The power of feedback in second language writing. *Journal of Second Language Writing*, 45, 1-13.

Pallant, J. (2013). *SPSS survival manual: A step-by-step guide to data analysis using IBM SPSS* (6th ed.). Open University Press.

Saldaña, J. (2016). The coding manual for qualitative researchers (3rd ed.). SAGE Publications.

Stevenson, M., & Phakiti, A. (2014). The effects of computer-generated feedback on EFL learners' writing accuracy. *Computer Assisted Language Learning*, 27(3), 248-262.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Wang, Z. (2020). Computer-assisted EFL writing and evaluations based on artificial intelligence: A case from a college reading and writing course. *Library Hi Tech*, 38(1), 172-193.

Ware, P., & Warschauer, M. (2006). Electronic feedback and second language writing. In K. Hyland & F. Hyland (Eds.), *Feedback in second language writing: Contexts and issues* (pp. 105-122). Cambridge University Press.

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). A 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.