International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 23s, 2025
https://theaspd.com/index.php

A Comparative Study Of AI-Generated Corrective Feedback And Teacher-Written Corrective Feedback Among A2-Level EFL Learners In Tertiary Education

Nilüfer Evişen[1*] **& Emrah Cinkara**[2]

[1] School of Foreign Languages, Gaziantep University, Gaziantep, Türkiye. Email: evisen@gantep.edu.tr
[2] Department of English Language Teaching, Gaziantep University, Gaziantep, Türkiye. Email: cinkara@gantep.edu.tr
[*] Corresponding author: evisen@gantep.edu.tr

Abstract: *This study compared teacher-written and AI-generated written corrective feedback on A2-level English-as-a-foreign-language (EFL) paragraphs to examine coverage, overlap, and pedagogical value. Learner texts (N = 39 paragraphs) received feedback from an experienced teacher and ChatGPT-4 (June 2025 release). Feedback was coded at the category level (Grammar, Vocabulary, Spelling & Punctuation, Syntax/word order) as present/absent for each learner × category pair. Paired analyses showed that ChatGPT flagged more feedback categories per paragraph than the teacher, with a significant within-pair effect (Wilcoxon signed-rank, p <.001; d = 1.15). A McNemar exact test on discordant pairs indicated a significant asymmetry favoring ChatGPT (p <.001), demonstrating that the model contributed substantially more unique category notices beyond those offered by the teacher. Category-specific contrasts revealed that the surplus was concentrated in Syntax and Vocabulary, whereas Grammar and Spelling & Punctuation exhibited near-complete overlap between rater types. The results suggest that large-language-model feedback can broaden the lexical–syntactic feedback net without sacrificing surface accuracy, offering efficiency gains while maintaining complementarity with teacher expertise. Pedagogically, the findings support hybrid workflows in which teachers curate or mediate AI suggestions to preserve motivational tone and help learners prioritize revisions. Limitations include focusing on a single proficiency band and binary category coding; future work should track uptake and learning gains and assess cost–benefit profiles of teacher-mediated AI feedback.*

Keywords: *AI in language education, A2 proficiency, ChatGPT, EFL writing, feedback overlap, written corrective feedback.*

1. INTRODUCTION

Despite the unstable question of whether feedback on second language learners' written products hones their writing skills or not [1], [2], many studies are available that emphasize the benefits that teacher written feedback can provide in the language acquisition process of L2 learners [3]– [6].

Teacher-written feedback is the notes, codes, symbols, and comments teachers provide on students' written assignments to improve their learning. It undertakes an important role in increasing the quality of students' written accounts. Through written feedback, students might have the chance to see their errors and/ or areas to be improved, and thus, become more proficient in expressing themselves via writing, thinking critically, and fostering their autonomy as learners [7].

Feedback is a valuable tool in learning, offering insights into students' progress and informing educators about their achievements and challenges [8]. It provides information on the performance of learning tasks, typically aimed at enhancing that performance [9], guiding both students and teachers in achieving educational goals [10]. Considering that learning is a social experience and thus necessitates interaction

between the teacher and the student, feedback is an element that is crucial for learning to take place on the students' side [11].

From the traditional aspect, feedback serves as a way of assessing learners' journey through the learning experience and teachers' achievement in assisting students throughout the process. Teachers provide guidance for their students in a formative way, focusing on the developmental period students are going through [12], and in a corrective way, focusing more on their errors. Over time, rigid writing activities with predefined boundaries, which prevent going beyond the set framework, have gradually been replaced by more interactive ones, and the benefit of the feedback provided by teachers for their students during this process is undeniable [13]. However, on the teachers' side, it should be noted that providing feedback for learners requires a lot of time and dedication.

Thus, the arrival of artificial intelligence to assist teachers in the task of examining and assessing students' written products seemed to be as timely as possible regarding the heavy workload and tight schedules teachers have and the contributions it provides for the students, namely, receiving prompt rather than delayed feedback [14], [15], [16]. Artificial intelligence has been reported to decrease teachers' stress levels by lightening their workload and equipping them with sufficient support [17], [18]. Few of the studies available on AI and its integration into education highlight that the use of AI tools such as Grammarly and ChatGPT may increase effective language learning as they provide automated writing evaluation (AWE) [19],[20], [21]. Compared to AI tools, teacher feedback has the advantage of refining not only students' writing skills but also language skills in general, as teacher feedback might come along with the oral interaction, a talent gifted to human beings [22], [23].

However, there is a gap in the literature of studies comparing teacher feedback and AI feedback in L2 writing skills. Until AI tools came into use, feedback had been practiced in the L2 environment for a long time. However, since this education technology is now available, teachers and students have started benefiting from it. Nevertheless, there is a need for more research on teacher feedback versus feedback provided by AI tools, especially in the digital era. Teachers' views on how AI handles students' written products and whether they would be willing to employ it in their teaching in the future deserve to be elaborated on more. Having stated all these, the current study explores the differences, advantages, and/ or disadvantages between teacher feedback and AI feedback on students' written accounts.

Specifically, the following three research questions will shed light on our research:

1. What are the key similarities between teacher and AI-generated feedback in L2 writing?
2. What are the teachers' reflections regarding the accuracy, usefulness, and alignment of AI feedback compared to their own?
3. How do teachers envision their future teaching via combining AI tools and their feedback to enhance L2 writing instruction?

2. LITERATURE REVIEW

A. Overview of Feedback in L2 Writing

Research suggests that feedback in L2 writing has numerous gains for learners in that it enhances students' realizing their language areas that need to be improved, and their knowledge on L2 [13] , [22], [23]; yet students' differences, prior education, sources of external and internal distraction, and motivation should not be left out of the picture [24], [25].The effectiveness of feedback on L2 writing might depend on cultural, institutional, and interpersonal factors [22] or ecological factors stemming from the learning environment, level of motivation, and learners' capacity [26]. However, it still serves as a stable tool for

interaction between both parties. Since writing is one of the most advanced skills to be mastered, becoming self-sufficient in an L2 writing context deserves more attention and patience on both the teachers' side and the learners [27]. It is among the teachers' tasks to point out mistakes and/ or areas to be improved regarding students' writing skills and then communicate this to their learners, which is just a first step in the long journey of learning a foreign language.

Many studies highlight the teacher's role in giving feedback [28]. To illustrate, Hattie and Timperley [29] and Wisniewski et al. [30] proffer that constructive and actionable feedback significantly affects student engagement, motivation, and achievement. Evans [31] states that effective feedback leads to and is crucial for learners' cognitive and behavioral development by providing room for self-reflection on the learners' side. The contribution of feedback regarding enhancing self-regulated learning, students' motivation levels, and overall achievement has also been articulated by Carless [32] and Yang et al. [33]. Wang et al. [34] investigated how effective teacher feedback was on students' English achievement in an online EFL classroom. Their study depicted that teacher feedback impacted students' English achievement directly and positively.

Teacher Feedback: Teachers have been providing feedback for students in various forms. The most common type of feedback is direct feedback, by which the teacher explicitly writes the corrected version of the part that contains the error [35]. Unlike direct feedback, teachers sometimes prefer to signal that the students have made a mistake by underlining, circling, coloring, and/ or putting a question mark next to the part that contains the error. This type of indirect feedback might be where the teacher underlines the part with the error without providing the corrected version [36]. The next type of written corrective feedback is metalinguistic feedback. Here, the teacher makes use of codes that symbolize the error. Metalinguistic feedback can also be a brief grammatical explanation within the margin or at the end of the text [3]. Teacher-written corrective feedback could also be classified further based on the focus of the feedback. Teachers might choose to point out each or all of the mistakes that the students have committed, unfocused corrective feedback, or they might prefer to highlight only specific errors, namely a single error type, which has been classified as focused corrective feedback [35].

Studies indicate that written corrective feedback significantly contributes to learners, regardless of the type of feedback provided. To illustrate, detailed and complete written feedback has been reported as contributing to student learning as regards the use of prepositions and simple past tense in Bitchener et al.'s [36] study. In her explorative study, Bayram [37] also found that grammar mistakes tend to decrease after students receive direct corrective feedback, which suggests that such feedback can address specific language errors. In a similar context, Nguyen [38] concluded that Vietnamese EFL learners had a preference for and a positive attitude towards written corrective feedback in a university context. Similarly, Zia et al. [39] contend that such feedback is a bridge to improving students' writing skills, leading to higher grades through formative assessments. Bonsu's study, carried out in 2021 [40], further supports the view that receiving corrective feedback has been appreciated and valued on both the students' and teachers' sides, as clear positive outcomes are reflected in students' writing abilities.

AI-generated Feedback: Instructors' perspectives regarding AI-generated feedback in L2 writing have been underrepresented [41]; however, the available studies explore the concept through the lens of learners and their insights on the effectiveness of such feedback. Kurt's study [42], for example, explored the perceptions of Turkish pre-service English teachers on ChatGPT as a feedback tool. It was seen that participants appreciated ChatGPT due to its practical, interactive, and adaptable feedback. However, some participants noted some inconsistencies with the quality of the prompts as well. Similarly, Sasaki et al. [43] compared AI-generated translation feedback with teacher corrective feedback among Japanese university students.

According to the findings of the study, AI-generated translation feedback assisted students with mastery in grammar and accuracy, while teacher feedback improved writing complexity. It could be concluded that AI-generated feedback was beneficial; yet there is still a need to conduct further research into instructors' views and the overall impact on L2 writing development.

In another study, Rahman et al. [44] focused on diagnosing and treating grammatical errors through an AI tool to develop EFL learners' writing skills. The study found that learners welcomed the technology as it contributed to their writing skills. Similarly, Marzuki et al. [45] examined the effects of AI-assisted language learning on EFL learners' writing performance. They concluded that AI technology assisted students in improving their writing skills compared to students who did not use AI tools.

3. METHODS

A. Participants and Data

The dataset comprised 39 learner-produced paragraphs, each written by A2-level learners of English as a foreign language. At this proficiency level, writers are typically able to produce short texts but still face challenges with accuracy, fluency, and structural control. Each learner text was paired with two sets of evaluative feedback: one provided by an experienced EFL teacher and the other generated by ChatGPT-4 (June 2025 release). This dual-source design enabled a direct comparison of human and machine feedback on the same learner productions.

B. Feedback Collection and Coding

Feedback from both raters was coded in a binary format ("flagged" vs. "not flagged") across four error categories that are widely documented in beginner writing research:

- Grammar (morphological and syntactic accuracy, e.g., tense marking, subject–verb agreement)
- Vocabulary (lexical choice, appropriateness, and variety)
- Spelling and Punctuation (orthographic accuracy and mechanical correctness)
- Syntax/Word Order (sentence structure and sequencing of elements)

This categorical framework ensured comparability between rater outputs while capturing surface-level and higher-order linguistic features. Each paragraph was thus coded across four dimensions, yielding 156 paired observations (39 paragraphs × 4 categories).

C. Analysis Procedures

We first computed descriptive statistics of category coverage for each rater type to evaluate rater concordance. Overlap and divergence between teacher and ChatGPT feedback were visualized using a Venn diagram to illustrate the distribution of shared and unique category-level judgments. To examine whether ChatGPT systematically flagged more categories than teachers, we compared the number of categories identified per paragraph using a Wilcoxon signed-rank test, a non-parametric test suitable for paired data. Effect size was calculated as d_z, allowing the magnitude of within-pair differences to be interpreted.

Mcnemar's exact tests were conducted separately for each error category to explore areas of agreement and disagreement further. This allowed us to assess whether discrepancies between teacher-only and chatgpt-only judgments reached statistical significance. Results are reported with exact p values, alongside descriptive summaries of the distribution of judgments across categories.

4. FINDINGS

B. Dataset Overview

The final dataset consisted of 39 learner-produced paragraphs, all written at the A2 level of the CEFR, which is typically associated with early-stage foreign language learners who can produce short texts but still struggle with accuracy and complexity. Each paragraph was accompanied by two independent sets of feedback: one generated by an experienced EFL teacher and the other produced by ChatGPT-4 (June 2025 release). This dual-source design ensured that every learner text was evaluated from a human and a machine perspective, allowing for direct comparison. To make the analyses systematic, all feedback was coded in a binary fashion ("flagged / not flagged") across four error-type categories well-documented in beginner writing research: grammar, vocabulary, spelling and punctuation, and syntax/word order. This coding scheme enabled a straightforward alignment between rater outputs while capturing the range of feedback tendencies. Multiplying the 39 learner texts by the four coding categories produced 156 paired observations, providing a balanced dataset for examining concordance and divergence between teacher and model judgments.

C. Volume of Feedback per Learner

Across the dataset, ChatGPT consistently identified more feedback categories per learner text than the human teacher. On average, the model flagged 3.76 categories (SD = 0.55), whereas teachers flagged an average of 2.76 categories (SD = 0.88). This indicates that, paragraph by paragraph, ChatGPT typically identified at least one additional issue beyond what the teacher recorded. To test whether this difference was statistically robust, we conducted a Wilcoxon signed-rank test, confirming that the effect was significant and reliable, V = 424.5, p < .001. The within-pair effect size was d_z = 1.15, which falls in the extensive range, underscoring the substantive nature of this difference. These results suggest that ChatGPT provides systematically broader coverage than teachers, a pattern unlikely to be due to chance variation.

D. Overlap of Feedback Categories

We conducted a category-level comparison across the entire corpus to provide a clearer picture of the results. The analysis focused on identifying areas of overlap and divergence between categories, highlighting both shared and unique features. The findings suggest that while several categories demonstrate substantial concordance, others remain distinct, reflecting nuanced differences in the dataset. The Venn diagram (Figure 1) visualizes this category-level concordance, illustrating the extent of overlap and separation among the groups in a single, accessible representation.

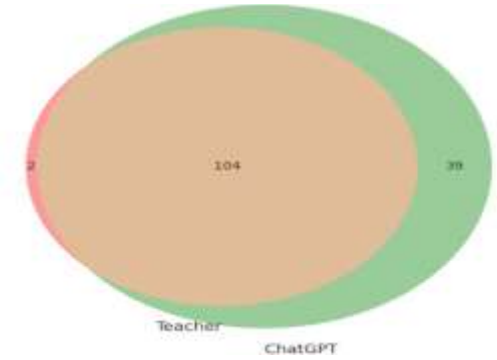

Figure 1. Category level concordances

The two raters converged on 72 % of all flagged categories (104 / 145 non-empty cases). A McNemar test comparing paired category flags showed a significant asymmetry, χ^2 (1) = 34.23, p <.001, with ChatGPT

adding 39 category notices that teachers missed versus only two teacher-only notices. A follow-up Wilcoxon signed-rank test on the number of categories flagged per learner confirmed the effect ($V = 10.5$, $p < .001$, $d_z = 1.15$).

Because the same learner/text forms the unit of comparison, a McNemar exact test was applied to the discordant pairs (Teacher-only = 2; ChatGPT-only = 39). The asymmetry was highly significant, $\chi^2(1) = 31.61$, $p < .001$ (exact $p \approx 7.8 \times 10^{-10}$). Hence, ChatGPT contributed a statistically greater number of additional category notices than the teacher supplied.

E. Category-Specific Patterns

Breaking down the McNemar contrasts by category revealed a differentiated profile highlighting consistency and discrepancies across the data. This analysis allowed us to pinpoint where significant shifts occurred and where stability was maintained. The results indicate that some categories showed strong agreement, while others diverged, suggesting a more complex underlying pattern. These differentiated outcomes provide a nuanced view of category-level dynamics, offering insight into how specific contrasts contribute to the findings.

Table 1. Rater Performance Across Categories

Category	Teacher-only	ChatGPT-only	Both	McNemar exact p
Grammar	1	1	37	.999
Spelling & Punctuation	1	1	34	.999
Vocabulary	0	**10**	24	.002
Syntax	0	**27**	9	< .001

Table 1 presents the distribution of rater performance across categories, comparing teacher-only judgments, ChatGPT-only judgments, and cases of agreement between the two. In the domains of grammar and spelling & punctuation, the high number of overlapping judgments (37 and 34, respectively) combined with non-significant McNemar exact p-values ($p = .999$ in both cases) indicates that teachers and ChatGPT reached almost identical conclusions, with virtually no meaningful divergence. By contrast, the vocabulary category revealed a notable imbalance: while there were no teacher-only ratings, ChatGPT produced ten unique judgments, leading to a statistically significant difference ($p = .002$). The most striking contrast emerged in the syntax category. ChatGPT generated twenty-seven unique ratings compared to none for teachers, leaving only nine shared judgments; the highly significant McNemar result ($p < .001$) underscores a clear divergence in evaluative focus. These results suggest that ChatGPT aligns closely with teacher judgments in more rule-governed areas such as grammar and mechanics but diverges substantially in higher-order linguistic dimensions like vocabulary and syntax, where its performance profile appears more expansive than human raters.

In conclusion, the analysis reveals a distinct coverage advantage, with ChatGPT reliably contributing at least one additional feedback category per learner paragraph. This represents a meaningful practical gain, as it ensures that student texts receive broader diagnostic attention than they would from teachers alone. Importantly, this additional coverage is not randomly distributed. Instead, it reflects a complementary focus, with ChatGPT's surplus comments clustering around higher-order concerns such as word order and lexical choice—areas that teachers tended to flag less frequently. This pattern suggests that the model may compensate for some of the natural economizing human raters engage in, thereby enriching the overall feedback profile. At the same time, the model shows no loss on basics: in the core domains of grammar and

spelling, where accurate error detection is critical, its coverage nearly mirrors that of teachers, with minimal discrepancies. These findings provide a robust quantitative baseline for evaluating the pedagogical value of large-language-model feedback. Specifically, they indicate that such systems can broaden the issues highlighted for novice writers while maintaining fidelity to fundamental error categories, thereby offering a potentially valuable complement to human instructional practices.

5. DISCUSSION AND CONCLUSION

The present study demonstrates that ChatGPT-4 provided a significantly broader range of feedback categories on A2-level learner paragraphs than an experienced EFL teacher. On average, the model identified one additional category per text, a difference that produced a large within-pair effect size ($d = 1.15$). McNemar's exact test confirmed the asymmetry in the discordant pairs of observations: ChatGPT offered 39 category notices that the teacher did not mention, whereas the teacher supplied only two unique notices. These quantitative results mirror earlier reports that large language models augment rather than merely replicate human feedback [46], [47]. Importantly, the surplus resided almost exclusively in vocabulary and syntax, while overlap with the teacher remained near-complete for grammar and mechanics.

Two complementary explanations help account for this pattern. First, transformer architectures process entire stretches of text through multi-head attention, enabling simultaneous detection of grammatical dependencies and lexical collocations that may escape a time-constrained teacher's notice. Eye-tracking research by [47] suggests that human raters devote disproportionate visual attention to surface-error hotspots, whereas ChatGPT evaluates cohesion and word order almost instantaneously. Second, the probabilistic nature of large language models encourages "hyper-vigilant" flagging of atypical but technically acceptable structures [48]. Teachers, in contrast, often downplay such grey-zone issues so as not to overload learners or undermine confidence. Together, these mechanisms position generative AI as a complementary lens that widens the lexical-syntactic feedback net without compromising core grammatical coverage.

Pedagogically, the findings imply that hybrid feedback systems may offer optimal benefit. ChatGPT produced the additional category notices in approximately 15 seconds per script, whereas the teacher required about six minutes, highlighting a substantial efficiency gain. Nevertheless, the model's feedback adopted a more directive tone, whereas the teacher's comments were hedged and motivational. Prior work shows learners perceive teacher-mediated AI feedback more positively than raw AI output [46]. Therefore, a workflow in which teachers curate or rephrase AI-generated comments could combine comprehensive coverage with affective sensitivity. Such a model frees teacher time for higher-order concerns such as idea development and genre-specific conventions.

Theoretically, the study contributes to noticing-based accounts of second-language writing. By surfacing additional lexical and syntactic issues, ChatGPT may elevate learners' awareness of patterns that conventional instruction leaves implicit. From a sociocultural perspective, however, mediation remains crucial; learners must negotiate and internalise the feedback through scaffolded dialogue. Accordingly, teacher expertise is central in helping students prioritise the AI's often voluminous suggestions.

Several limitations temper the generalisability of these conclusions. The dataset was restricted to A2-level writers; intermediate or advanced learners might display different overlap profiles. The binary coding scheme measured only whether a category was mentioned, not individual comments' precision or pedagogical usefulness. In addition, the findings pertain to the June 2025 release of ChatGPT-4; subsequent model iterations may alter feedback behaviour, underscoring the need for version-locked replication.

Future research should examine whether the AI-only category notices translate into successful revisions and long-term accuracy gains. Mixed-methods designs combining revision tracking with learner interviews could illuminate how students negotiate the cognitive load of expanded feedback. Cost–benefit analyses incorporating teacher time, subscription fees, and learning outcomes would further clarify the practical viability of integrating generative AI into formative assessment routines.

In sum, generative AI is not a replacement for human expertise but a potent complement. When embedded thoughtfully within teacher-mediated feedback cycles, models such as ChatGPT-4 can broaden the scope of issues addressed, especially in lexical and syntactic domains, while allowing educators to reallocate attention to higher-order writing concerns and affective support.

REFERENCES

1. Lee, "Understanding Teachers' Written Feedback Practices In Hong Kong Secondary Classrooms," J. Second Lang. Writ., Vol. 17, No. 2, Pp. 69–85, 2008.
2. J. Truscott, "The Case Against Grammar Correction In L2 Writing Classes," Lang. Learn., Vol. 46, No. 2, Pp. 327–369, 1996.
3. M. M. Albogami, "An Inquiry Into Effective Written Feedback From Efl Teachers' And Students' Perspectives At A Saudi University," Ph.D. Dissertation, Univ. Exeter, Exeter, U.K., 2020.
4. Ç. Atmaca, "The Role Of Written Corrective Feedback In Developing Academic Writing," J. Lang. Linguist. Stud., Vol. 12, No. 2, Pp. 1–14, 2016.
5. K. Hyland And F. Hyland, "Feedback On Second Language Students' Writing," Lang. Teach., Vol. 39, No. 2, Pp. 83–101, 2006.
6. N. Kaya, "Written Corrective Feedback In Efl Writing: Teachers' Beliefs And Practices," Int. Online J. Educ. Teach., Vol. 10, No. 1, Pp. 125–140, 2023.
7. M. A. Adarkwah, "The Power Of Assessment Feedback In Teaching And Learning: A Narrative Review And Synthesis Of The Literature," Sn Social Sci., Vol. 1, No. 3, P. 75, 2021.
8. Audras, A. Zhao, C. Isgar, And Y. Tang, "Virtual Teaching Assistants: A Survey Of A Novel Teaching Technology," Int. J. Chin. Educ., Vol. 11, No. 2, 2022, Art. No. 2212585x221121674.
9. R. Hashem Et Al., "Ai To The Rescue: Exploring The Potential Of Chatgpt As A Teacher Assistant To Reduce Workload And Prevent Burnout In Secondary Schools," Res. Pract. Technol. Enhanc. Learn., Vol. 19, Art. No. 23, 2024.
10. J. Roberts, "Ai Writing Feedback For Students," Edutopia, Oct. 30, 2024.
11. J. Escalante, A. Pack, And A. Barrett, "Ai-Generated Feedback On Writing: Insights Into Efficacy And Enl Student Preference," Int. J. Educ. Technol. Higher Educ., Vol. 20, No. 1, P. 57, 2023.
12. S. J. Ingley And A. Pack, "Leveraging Ai Tools To Develop The Writer Rather Than The Writing," Comput. Educ. Ai, 2023.
13. J. A. Boggs And R. M. Manchón, "Feedback Literacy In Writing Research And Teaching: Advancing L2 Wcf Research Agendas," Assess. Writ., Vol. 58, P. 100786, 2023.
14. R. Ferris, "Written Corrective Feedback For Individual L2 Writers: A Longitudinal Multiple-Case Study," J. Second Lang. Writ., Vol. 22, No. 2, Pp. 1–22, 2013.
15. K. Hyland And F. Hyland, *Feedback In Second Language Writing: Contexts And Issues*, 2nd Ed., Cambridge, U.K.: Cambridge Univ. Press, 2019.

16. Z. Zhang And K. Hyland, "Student Engagement With Peer Feedback In L2 Writing: Insights From Reflective Journaling And Revising Practices," Assess. Writ., Vol. 58, Art. No. 100784, Oct. 2023.
17. Y. Han, "Written Corrective Feedback From An Ecological Perspective: The Interaction Between The Context And Individual Learners," System, Vol. 80, Pp. 288–303, 2019.
18. B. M. Wondim, K. S. Bishaw, And Y. T. Zeleke, "Effectiveness Of Teachers' Direct And Indirect Written Corrective Feedback Provision Strategies On Enhancing Students' Writing Achievement: Ethiopian University Entrants In Focus," Heliyon, Vol. 10, No. 2, 2024.
19. S. Yu, "Feedback-Giving Practice For L2 Writing Teachers: Friend Or Foe?," J. Second Lang. Writ., Vol. 52, P. 100798, 2021.
20. J. Hattie And H. Timperley, "The Power Of Feedback," Rev. Educ. Res., Vol. 77, No. 1, Pp. 81–112, 2007.
21. B. Wisniewski, J. Zierer, And J. Hattie, "The Power Of Feedback Revisited: A Meta-Analysis Of Educational Feedback Research," Front. Psychol., Vol. 10, Art. No. 3087, 2019.
22. C. Evans, "Making Sense Of Assessment Feedback In Higher Education," Rev. Educ. Res., Vol. 83, No. 1, Pp. 70–120, 2013.
23. Carless, *Excellence In University Assessment: Learning From Award-Winning Practice*. Abingdon, U.K.: Routledge, 2015.
24. M. Yang, D. Carless, And D. Lau, "Student Self-Evaluation: Teachers' Perspectives In Hong Kong," Assess. Educ., Vol. 29, No. 2, Pp. 125–141, 2022.
25. Y. Wang, C. Shang, And R. Briody, "The Impact Of Teacher Feedback On Students' English Achievement In Online Efl Classrooms," Int. J. Educ. Technol. Higher Educ., Vol. 19, Art. No. 12, 2022.
26. R. Ellis, Y. Sheen, M. Murakami, And H. Takashima, "The Effects Of Focused And Unfocused Written Corrective Feedback In An English As A Foreign Language Context," System, Vol. 36, No. 3, Pp. 353–371, 2008.
27. J. Bitchener, S. Young, And D. Cameron, "The Effect Of Different Types Of Corrective Feedback On Esl Student Writing," J. Second Lang. Writ., Vol. 14, No. 3, Pp. 191–205, 2005.
28. Bayram, "The Effects Of Direct Written Corrective Feedback On Turkish Efl Students' Writing Accuracy," Int. J. Lang. Educ., Vol. 2, No. 1, Pp. 1–15, 2016.
29. T. Nguyen, "The Role Of Written Corrective Feedback In Improving Vietnamese Efl Learners' Writing Performance," J. Lang. Teach. Res., Vol. 15, No. 2, Pp. 200–212, 2024.
30. Zia, S. Sarfraz, And N. Mufti, "Students' Perceptions Of The Effectiveness Of Formative Assessment And Feedback For Improvement Of The English Writing Composition Skills: A Case Study Of Secondary Level Esl Students Of Private Schools In Lahore, Pakistan," J. Educ. Pract., Vol. 10, No. 6, Pp. 7–13, 2019.
31. Bonsu, "The Influence Of Written Feedback On The Writing Skill Performance Of High School Students," Int. J. Appl. Res. Social Sci., Vol. 3, No. 3, Pp. 33–43, 2021.
32. D. Wang, "Teacher-Versus Ai-Generated (Poe Application) Corrective Feedback And Language Learners' Writing Anxiety, Complexity, Fluency, And Accuracy," Int. Rev. Res. Open Distrib. Learn., Vol. 25, No. 3, Pp. 37–56, 2024.
33. Kurt, "Exploring Pre-Service English Teachers' Perceptions Of Chatgpt As A Feedback Tool," Int. J. Educ. Technol. Learn., Vol. 9, No. 1, Pp. 12–25, 2024.
34. M. Sasaki, Y. Sato, And H. Takahashi, "Comparing Ai-Generated Translation Feedback And Teacher Corrective Feedback In Japanese University Efl Writing Classes," Asian Efl J., Vol. 28, No. 3, Pp. 44–63, 2024.

35. M. M. Rahman, S. M. Islam, And T. Ferdous, "Using Artificial Intelligence To Diagnose And Treat Grammatical Errors In Efl Writing: Implications For Language Learning," Comput. Assist. Lang. Learn., Vol. 35, No. 7, Pp. 1456–1474, 2022.
36. M. Marzuki, R. Wulandari, And H. Ismail, "Ai-Assisted Language Learning: Improving Efl Learners' Writing Performance," Arab World Engl. J., Vol. 14, No. 1, Pp. 123–140, 2023.
37. Alnemrat, R. A. Al-Shboul, And L. Al-Saleem, "Large Language Models In Education: Augmenting Teacher Feedback In Efl Writing Instruction," Educ. Inf. Technol., Vol. 30, No. 2, Pp. 2341–2360, 2025.
38. J. Han And M. Li, "Augmenting Teacher Feedback With Ai: A Study Of Llms In Efl Writing," System, Vol. 123, P. 103567, 2024.
39. L. Hirschi, "The Hyper-Vigilance Of Large Language Models: Acceptability Judgments In Ai-Generated Feedback," Comput. Assist. Lang. Learn., Vol. 37, No. 4, Pp. 789–807, 2024.
40. Bonsu, "The Influence Of Written Feedback On The Writing Skill Performance Of High School Students," Int. J. Appl. Res. Social Sci., Vol. 3, No. 3, Pp. 33–43, 2021.
41. D. Wang, "Teacher-Versus Ai-Generated (Poe Application) Corrective Feedback And Language Learners' Writing Anxiety, Complexity, Fluency, And Accuracy," Int. Rev. Res. Open Distrib. Learn., Vol. 25, No. 3, Pp. 37–56, 2024.
42. Kurt, "Exploring Pre-Service English Teachers' Perceptions Of Chatgpt As A Feedback Tool," Int. J. Educ. Technol. Learn., Vol. 9, No. 1, Pp. 12–25, 2024.
43. M. Sasaki, Y. Sato, And H. Takahashi, "Comparing Ai-Generated Translation Feedback And Teacher Corrective Feedback In Japanese University Efl Writing Classes," Asian Efl J., Vol. 28, No. 3, Pp. 44–63, 2024.
44. M. M. Rahman, S. M. Islam, And T. Ferdous, "Using Artificial Intelligence To Diagnose And Treat Grammatical Errors In Efl Writing: Implications For Language Learning," Comput. Assist. Lang. Learn., Vol. 35, No. 7, Pp. 1456–1474, 2022.
45. M. Marzuki, R. Wulandari, And H. Ismail, "Ai-Assisted Language Learning: Improving Efl Learners' Writing Performance," Arab World Engl. J., Vol. 14, No. 1, Pp. 123–140, 2023.
46. Alnemrat, R. A. Al-Shboul, And L. Al-Saleem, "Large Language Models In Education: Augmenting Teacher Feedback In Efl Writing Instruction," Educ. Inf. Technol., Vol. 30, No. 2, Pp. 2341–2360, 2025.
47. J. Han And M. Li, "Augmenting Teacher Feedback With Ai: A Study Of Llms In Efl Writing," System, Vol. 123, P. 103567, 2024.
48. Hirschi, "The Hyper-Vigilance Of Large Language Models: Acceptability Judgments In Ai-Generated Feedback," Comput. Assist. Lang. Learn., Vol. 37, No. 4, Pp. 789–807, 2024.