

# AI-Assisted Feedback versus Teacher Feedback in Improving EFL Learners' Writing Accuracy

Hongxia Li<sup>1</sup>, Xin Tuo<sup>2</sup>

<sup>1</sup>*Xi'an Innovation College of Yan'an University, Xi'an, China*

<sup>2</sup>*Xi'an Innovation College of Yan'an University, Xi'an, China*

**Abstract:** *This mixed-methods study compares AI-assisted feedback and teacher feedback in improving EFL learners' writing accuracy. Sixty intermediate Chinese university EFL learners were assigned to an AI group and a Teacher group. Data came from pre-/post-tests, error analysis, interviews, and journals. Both feedback types significantly improved overall accuracy, with no significant group difference. AI feedback was stronger for surface-level errors, whereas teacher feedback was stronger for deep-level errors such as coherence and word choice. Learners valued AI for immediacy but preferred teachers for contextual relevance, emotional support, and complex explanation. The study supports a blended feedback model guided by AI literacy and teacher mediation.*

**Keywords:** *AI-Assisted Feedback, Teacher Feedback, EFL Writing, Writing Accuracy, Written Corrective Feedback, Mixed-Methods Study*

## 1. Introduction

Writing accuracy is central to EFL writing proficiency, involving grammar, vocabulary, punctuation, and discourse control [1]. Learners often struggle because of L1 interference, limited linguistic knowledge, and insufficient practice [2]. WCF helps them notice errors, understand causes, and revise more effectively [3].

Teacher feedback remains dominant because it offers contextualized, personalized guidance [4]. Yet in large EFL classes, detailed feedback is time-consuming, often delayed, and vulnerable to teacher fatigue [5][6]. Delay can weaken uptake because learners lose revision momentum.

Generative AI has changed WCF by providing immediate support. ChatGPT and Grammarly can identify surface errors, offer explanations, and sometimes comment on coherence and word choice [7][8][9], reducing workload and supporting timely revision.

The comparative effectiveness of AI and teacher feedback remains debated. Some studies report comparable gains from AI feedback [8][10], while others emphasize teachers' superiority for contextual errors [9][11]. Few studies combine outcomes, perceptions, and uptake in non-Western EFL contexts.

This study compares AI-assisted feedback and teacher feedback in improving writing accuracy, examines learner perceptions, and identifies factors shaping feedback uptake. It addresses three research questions:

- 1) Is there a significant difference in the improvement of EFL learners' writing accuracy between those receiving AI-assisted feedback and those receiving teacher feedback?
- 2) Do AI-assisted feedback and teacher feedback differ in their effectiveness in correcting surface-level and deep-level writing errors?
- 3) What are EFL learners' perceptions of AI-assisted feedback and teacher feedback, and how do these perceptions influence feedback uptake?

The study contributes empirical evidence on the relative strengths of AI and teacher feedback and offers practical guidance for integrating AI tools into EFL writing instruction.

## 2. Literature Review

### 2.1. *Writing Accuracy and Written Corrective Feedback in EFL Writing*

Writing accuracy includes surface-level accuracy, such as grammar, punctuation, and spelling, and deep-level accuracy, such as word choice, coherence, cohesion, and genre appropriateness [1][4]. Both dimensions matter because errors obstruct meaning and credibility.

WCF helps learners notice, correct, and avoid recurring errors [1][3]. It may be direct, indirect, or metalinguistic, and its effectiveness depends on feedback type, error type, proficiency, and uptake [11][12].

### 2.2. *Teacher Feedback in EFL Writing*

Teacher feedback is valued for contextual sensitivity, personalization, and metalinguistic explanation. Teachers can consider learner proficiency, task purpose, and intended meaning, while also providing encouragement that supports motivation and uptake [3][4][12].

Its limitations are equally clear. Large classes make timely and detailed feedback difficult, and teachers may vary in focus or correction strategy [5][8]. Heavy workloads can also produce feedback fatigue and reduce quality [6][13].

### 2.3. *AI-Assisted Feedback in EFL Writing*

AI-assisted feedback uses LLMs and AWE systems to provide instant WCF [7][14]. AI tools are especially useful for grammar, punctuation, spelling, and some word-choice errors, while advanced tools can suggest coherence and cohesion improvements [8][9].

The main advantages of AI feedback are immediacy, scalability, and consistency. Learners can revise while the writing process is fresh, and teachers can shift repetitive correction work to digital tools [6][13][15].

AI feedback also has limits. It may miss context-dependent meanings, idiomatic use, cultural references, and genre conventions requiring human judgment [9]. It can overemphasize surface correction and may encourage over-reliance if learners accept suggestions without reflection [13][16].

### 2.4. *Comparative Studies of AI-Assisted Feedback and Teacher Feedback*

Comparative findings are mixed. Some studies report that AI feedback is comparable to teacher feedback and may reduce anxiety [8][10]. Others find teacher feedback stronger for deep-level accuracy because teachers provide contextual explanation and emotional support [9][11].

These findings suggest that AI and teacher feedback may not be competing alternatives but complementary resources. AI appears strongest for frequent mechanical errors, while teachers remain essential for meaning, discourse, and motivation.

A key gap is the shortage of mixed-methods studies combining accuracy outcomes with learner perceptions and feedback uptake. This study addresses that gap by examining both objective effectiveness and subjective experience.

## 3. Methodology

A mixed-methods quasi-experimental design was used to compare AI-assisted and teacher feedback. Quantitative data measured writing accuracy before and after the intervention, while interviews and journals explored learner perceptions and feedback uptake [17].

### 3.1. *Research Context and Participants*

The study involved 60 intermediate EFL learners at a Chinese university (32 female, 28 male; ages 19-22). They were randomly assigned to an AI group ( $n = 30$ ) and a Teacher group ( $n = 30$ ), with comparable proficiency and pre-test accuracy ( $t = 0.87$ ,  $p = .39$ ).

Participants gave written consent and were assured anonymity. Two experienced EFL teachers supported the study: one provided teacher feedback and the other trained the AI group in responsible tool use.

### **3.2. Research Instruments**

#### **1) Pre-Test and Post-Test**

Pre- and post-tests required 45-minute argumentative essays. Two raters scored accuracy with a modified ESL Composition Profile covering surface- and deep-level errors. Reliability was high (Cohen's kappa = 0.85), and disagreements were resolved through discussion.

#### **2) Error Analysis**

Error analysis classified errors as surface-level (grammar, punctuation, spelling, capitalization) or deep-level (word choice, collocation, coherence, cohesion, genre appropriateness). Errors per 100 words and error reduction rates were calculated for each participant.

#### **3) Semi-Structured Interviews**

Semi-structured interviews were conducted with 20 participants, 10 from each group, after the intervention. Interviews explored perceived effectiveness, strengths and limits of feedback, feedback uptake, and feedback preference.

#### **4) Reflective Journals**

Participants kept weekly reflective journals for 10 weeks, recording received feedback, revision choices, perceived accuracy changes, confidence, and difficulties. In total, 600 journal entries were collected.

### **3.3. Intervention Procedure**

The intervention lasted 10 weeks, with one argumentative writing task per week. Tasks, writing time, and materials were kept consistent across groups to support internal validity.

AI group participants used ChatGPT (GPT-4) and Grammarly. After initial training in prompting, interpreting feedback, and integrating suggestions, they submitted drafts to both tools, requested explanations when needed, and revised accordingly.

Teacher group participants submitted drafts to an experienced EFL teacher, who provided direct correction for surface errors, indirect correction for deep-level errors, and metalinguistic explanations when necessary. Feedback was returned within 48 hours.

### **3.4. Data Analysis**

#### **1) Quantitative Data Analysis**

Quantitative data were analyzed in SPSS 26.0. Paired-sample t-tests measured within-group improvement, independent-sample t-tests compared post-test scores and error reduction rates, and Cohen's d indicated effect size.

#### **2) Qualitative Data Analysis**

Interview and journal data were analyzed thematically following Braun and Clarke [18]. Two researchers coded independently, with high reliability (Cohen's kappa = 0.82); disagreements were discussed, and member checking was conducted with 10 participants.

### **3.5. Ethical Considerations**

The study followed AERA ethical guidelines [19]. Data were stored securely, participants could withdraw without penalty, and the AI group received guidance on avoiding over-reliance and ensuring that writing remained their own work.

#### 4. Results

Results are reported through quantitative comparisons of writing accuracy and qualitative analysis of learner perceptions and feedback uptake.

##### 4.1. Quantitative Results: Effectiveness in Improving Writing Accuracy

###### 1) Overall Writing Accuracy Improvement

Both groups improved significantly. The AI group's mean score rose from 62.33 to 75.67 ( $t = -12.45$ ,  $p < .001$ ,  $d = 2.63$ ), and the Teacher group's from 61.87 to 74.93 ( $t = -11.89$ ,  $p < .001$ ,  $d = 2.47$ ). The post-test difference was not significant ( $t = 0.58$ ,  $p = .56$ ).

###### 2) Error Correction Effectiveness by Error Type

Error analysis showed different strengths by error type (Table 1).

*Table 1: Error Reduction Rates (%) for Surface-Level and Deep-Level Errors.*

| Feedback type        | Surface-level errors | Deep-level errors |
|----------------------|----------------------|-------------------|
| AI-assisted feedback | 68.32 (SD = 7.45)    | 42.15 (SD = 8.12) |
| Teacher feedback     | 59.47 (SD = 8.03)    | 58.63 (SD = 7.68) |
| t-value              | 3.82                 | -6.25             |
| p-value              | < .001               | < .001            |
| Cohen's d            | 1.12                 | -1.85             |

The AI group reduced surface-level errors more than the Teacher group ( $t = 3.82$ ,  $p < .001$ ,  $d = 1.12$ ), showing strength in grammar, punctuation, spelling, and capitalization. The Teacher group reduced deep-level errors more effectively ( $t = -6.25$ ,  $p < .001$ ,  $d = -1.85$ ), especially word choice, collocation, coherence, and cohesion.

##### 4.2. Qualitative Results: Learners' Perceptions and Feedback Uptake

Thematic analysis generated three themes: AI feedback's strengths and limitations, teacher feedback's strengths and limitations, and factors shaping feedback uptake.

###### 1) Strengths and Limitations of AI-Assisted Feedback

AI group participants highlighted immediacy and convenience. Instant feedback allowed them to revise before losing track of their ideas:

ChatGPT and Grammarly give me feedback right away, so I can fix my grammar and spelling errors immediately. I don't have to wait for days like I do with teacher feedback. (Participant AI-7, University EFL Learner)

Journals showed that immediate AI feedback increased motivation to revise and improved writing efficiency.

Participants also valued the consistency and detail of AI feedback for surface-level errors. One learner wrote:

Grammarly not only tells me I made a grammar error but also explains why it's wrong and how to fix it. This helps me remember the rule and avoid making the same error again. (Participant AI-12, University EFL Learner)

At the same time, learners reported that AI feedback was sometimes generic or contextually inappropriate. Some corrections did not match intended meaning, and AI tools struggled with idioms, cultural references, and task-specific expression.

AI was also weaker for deep-level writing problems. Learners reported that it corrected grammar but gave limited help with logical flow, cohesion, and argument development, areas where they still wanted teacher guidance.

###### 2) Strengths and Limitations of Teacher Feedback

Teacher group participants emphasized contextual sensitivity and personalization. They felt the teacher understood their intended meaning, writing style, and task context, allowing feedback to address coherence and ideas rather than grammar alone.

Learners also valued teachers' metalinguistic explanations and emotional support. Teachers could explain complex errors accessibly while encouragement increased confidence and motivation:

My teacher's feedback is not just about correcting errors—she also tells me what I did well and encourages me to keep improving. This makes me more confident in my writing. (Participant T-3, University EFL Learner)

The main limitation of teacher feedback was delay. Waiting two to three days made revision less immediate and sometimes weakened motivation. Some learners also noted that feedback focus varied across tasks.

### 3) Factors Influencing Feedback Uptake

Two factors shaped feedback uptake: AI literacy and teacher guidance. AI group learners with stronger prompting, interpretation, and revision skills used feedback more effectively, while those with lower AI literacy struggled to apply general suggestions.

Across both groups, teacher guidance improved uptake by showing learners how to interpret comments, revise step by step, and avoid repeated errors. One Teacher group participant wrote:

My teacher showed me how to use her feedback to revise my essay step by step. This helped me understand how to correct my errors and improve my writing accuracy. (Participant T-7, University EFL Learner)

AI group learners also benefited from initial teacher training, which helped them use AI tools more strategically and critically.

## 5. Discussion

This study clarifies the comparative value of AI-assisted and teacher feedback by combining accuracy outcomes with learner perceptions. The findings support a complementary rather than replacement-based view of AI in EFL writing instruction.

Quantitative results show that both feedback types improved overall accuracy, with no significant group difference. AI was stronger for surface-level errors, while teacher feedback was stronger for deep-level errors. This supports claims that AI excels at mechanical correction but is less reliable for context-dependent issues requiring human judgment [9][11].

Qualitative findings explain this pattern. Learners valued AI feedback for immediacy, convenience, and consistency, but criticized its limited contextual sensitivity. Conversely, they valued teacher feedback for personalization, explanation, and emotional support, while noting that delayed return reduced revision momentum.

Feedback uptake emerged as a mediating factor. AI feedback was most useful when learners had enough AI literacy to prompt, evaluate, and adapt suggestions. Teacher guidance remained essential for helping learners use both AI- and teacher-provided feedback productively.

### 5.1. Theoretical Implications

The study extends WCF research by showing that AI tools can function as useful feedback agents, particularly for surface-level errors. It also supports an error-type perspective: feedback should be matched to whether errors are mechanical or context-dependent.

The findings also connect to Self-Determination Theory [20]. AI feedback may enhance autonomy through immediacy and control, while teacher feedback supports competence and relatedness through explanation and encouragement.

### 5.2. Practical Implications

#### 1) For EFL Teachers

Teachers should adopt a blended feedback model. AI can handle frequent surface-level correction, while teachers focus on coherence, word choice, argumentation, and learner support. Teachers should also teach AI literacy, help learners evaluate suggestions, and prevent over-reliance.

#### 2) For Curriculum Designers

Curricula should integrate AI feedback into process writing rather than treating it as an optional shortcut. Modules on AI literacy, feedback uptake, and revision strategies can help learners use both AI and teacher feedback effectively.

### 3) For Educational Institutions

Institutions should provide reliable AI tools, technical support, teacher professional development, and clear ethical guidelines for AI use in writing instruction.

## 5.3. Limitations and Future Research

The study is limited by its single-university context, intermediate-level sample, and focus on argumentative writing. Future studies should examine other educational contexts, proficiency levels, and genres.

The study also used only ChatGPT and Grammarly; future work could compare different AI tools and identify which are most effective for specific error types.

Finally, the 10-week duration limits conclusions about long-term development. Future research should track sustained accuracy gains and examine teachers' perceptions of AI-assisted feedback.

## 6. Conclusion

This study found that AI-assisted and teacher feedback both improved EFL learners' writing accuracy. AI was stronger for surface-level errors, while teacher feedback was stronger for deep-level errors. Learners valued AI for immediacy but preferred teachers for contextual relevance, explanation, and support. Uptake depended on AI literacy and teacher guidance.

AI feedback should not replace teacher feedback. A blended model can use AI for rapid mechanical correction and teachers for higher-order writing, ethical guidance, and motivation, making feedback more efficient while preserving human expertise.

As AI develops, EFL educators and institutions need to balance technological efficiency with pedagogical judgment so learners develop both accuracy and independent writing ability.

## References

- [1] Ferris, D. R. (2006). *Corrective feedback in second language writing*. University of Michigan Press.
- [2] Han, F., & Hyland, K. (2015). Written corrective feedback in L2 writing: A meta-analysis. *Journal of Second Language Writing*, 29, 1–18. <https://doi.org/10.1016/j.jslw.2015.06.004>
- [3] Lee, J. F. (2019). Written corrective feedback in L2 writing: Contexts, issues, and challenges. *Annual Review of Applied Linguistics*, 39, 265–286. <https://doi.org/10.1017/S0267190519000089>
- [4] Hyland, K. (2019). *Second language writing* (2nd ed.). Cambridge University Press.
- [5] Kim, Y. (2014). Teacher feedback in EFL writing classrooms: A case study of Korean secondary school teachers. *Journal of Language Teaching and Research*, 5(3), 587–594. <https://doi.org/10.4304/jltr.5.3.587-594>
- [6] Hyland, K. (2025). EFL teachers and feedback fatigue: AI to the rescue? *ELT Journal*, 79(2), 189–198. <https://doi.org/10.1093/elt/ccae052>
- [7] Kasneci, G., et al. (2023). Large language models in EFL education: Opportunities and challenges. *Computers & Education*, 201, 104678. <https://doi.org/10.1016/j.compedu.2023.104678>
- [8] Yu, H., Li, J., & Zhang, Y. (2025). Generative AI vs. teachers: Feedback quality, feedback uptake, and revision. *Language Teaching Research Quarterly*, 47(2), 113–137. <https://doi.org/10.32038/ltrq.2025.47.06>
- [9] Lim, J. M., & Go, C. (2025). Integrating ChatGPT into teacher feedback: Practical insights for L2 writing instruction. *TESL-EJ*, 29(3), 1–18.
- [10] Wang, D. (2024). Teacher- versus AI-generated (Poe application) corrective feedback and language learners' writing anxiety, complexity, fluency, and accuracy. *The International Review of Research in Open and Distributed Learning*, 25(3), 123–145. <https://doi.org/10.19173/irrodl.v25i3.7734>
- [11] Webb, S., & Şen Aydın, Z. (2025). Generative AI in English language teaching: Students' voices, teachers' reactions, and needs. U.S. Department of Education. <https://files.eric.ed.gov/fulltext/EJ1488670.pdf>
- [12] Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in*

motivation, development, and wellness. Guilford Press.

[13] Poh, S. K., & Lee, I. (2026). *Generative artificial intelligence (GenAI) for feedback in school writing: Friend or foe?* *TESL-EJ*, 29(4), 1–20.

[14] Mahmoud, A. (2025). *AI powered ELT: Instructors' transformative roles and opportunities*. *PLOS ONE*, 19(7), e0324910. <https://doi.org/10.1371/journal.pone.0324910>

[15] Joo, M., Lee, J., & Park, H. (2025). *The impact of AI feedback efficacy on English writing achievement*. *Journal of Applied Information Science*, 26(5), 456–472. <https://doi.org/10.1000/jais.2025.265.456>

[16] Kohnke, L., Zhang, L., & Doe, C. (2024). *The differential impact of AI tools among EFL university learners: A process writing approach*. *International Journal of Learning, Teaching and Educational Research*, 23(5), 123–142. <https://doi.org/10.26803/ijlter.23.5.7>

[17] Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). SAGE Publications.

[18] Braun, V., & Clarke, V. (2006). *Using thematic analysis in psychology*. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>

[19] AERA. (2018). *Ethical standards of the American Educational Research Association*. American Educational Research Association.

[20] Deci, E. L., & Ryan, R. M. (2000). *The “what” and “why” of goal pursuits: Human needs and the self-determination of behavior*. *Psychological Inquiry*, 11(4), 227–268. [https://doi.org/10.1207/S15327965PLI1104\\_01](https://doi.org/10.1207/S15327965PLI1104_01)