Qom Branch
Islamic Azad University

Research Article

Human, AI, and Combined Corrective Feedback in EFL Writing: A Mixed Methods Comparative Study with Iranian Learners

Kolsoum Ghasemi[1] ✉, Shahram Afraz[2], Maryam Habouti[3]

[1] Department of English, Hormozgan University of Medical Sciences, Bandar Abbas, Iran
[2] Department of English Language Teaching, Qe.C., Islamic Azad University, Qeshm, Iran
[3] Ministry of Education, Bandar Anzali, Iran

Abstract

Corrective Feedback is a critical factor in enhancing EFL students' writing ability. This research contrasted the impact of the AI-generated, human-created, and blended written corrective feedback on Iranian EFL university students' academic writing proficiency, focusing on both surface-level accuracy and higher-order writing skills, such as coherence and organization. 384 intermediate-level students were randomly distributed across three groups and provided with feedback on their weekly writing assignments over on six weeks. To that end, IELTS-aligned rubrics were used to measure their writing performance, and students' views were probed through focus group interviews. Based on the quantitative data analyses, all three groups improved statistically. In fact, the combined written corrective feedback condition was the best, with the largest gain score and largest effect size, and the AI-generated written corrective feedback also led to robust gains, particularly on expanding grammatical range and lexical accuracy, while the human-only written corrective feedback yielded moderate-to-large effects, particularly on coherence and idea development. All groups showed statistically significant improvement in their writing performances. However, the combined feedback group outperformed both AI-generated and human-only groups, with the highest gain score. Four major themes and seven subthemes were extracted based on the qualitative data analysis of the focus group interviews with 30 interviews, Thematic analysis revealed that AI-generated WCF enhanced efficiency and reduced anxiety, human-generated WCF provided deeper conceptual guidance, and the combined WCF led to greater clarity, confidence, and more effective revision strategies among EFL learners. The findings suggest additional research and practice to see if the long-term influence of feedback modality and efficacy for learners with different proficiency levels and backgrounds holds.

Keywords: AI feedback, automated writing evaluation, EFL learners, human corrective feedback, Iranian students, writing performance

1. Introduction

Writing proficiency is central to second-language acquisition, especially in EFL contexts where learners struggle with grammar, vocabulary, and coherence. In this regard, written corrective feedback (WCF) plays a pivotal role in guiding revision and improving writing quality (Tran, 2025). Human-generated WCF offers context-sensitive and thoughtful input, while AI-generated WCF provides real-time corrections. However, both approaches have limitations: AI feedback may be overly generic or lack nuance, whereas human feedback can be time-consuming and inconsistently applied in large classes (Bai & Nordin, 2025; Li et al., 2024; Wang, 2024).

Despite the effectiveness of each WCF type, the comparative and combined application of the AI- generated and human-generated WCF in EFL contexts remains underexplored. Most empirical studies have examined the teacher-generated and AI-generated WCF separately. For instance, Wang's (2024) quasi-experimental research showed that the AI-generated WCF significantly lowered writing anxiety compared to teacher-generated WCF. Similarly, Tran (2025) investigated sequencing effects (AI before human), noting their influence on learners' revisions without addressing broader pedagogical concerns like classroom practice or scalability. A few comparative studies, such as Cao and Zhong (2023), have tested targeted language tasks (e.g., translations), finding AI stronger in lexical accuracy while humans surpass in syntactic complexity. Comprehensive reviews have also highlighted the need for more integrative research focusing on global writing quality and learner affective factors, including motivation and anxiety (Karagöz, 2025; Lee e al., 2025; Pratama & Sulistiyo, 2024; Shi & Aryadoust, 2024).

Filling this gap carries both educational and theoretical significance. Practically, combining the AI- and human-generated WCF maximizes their complementary strengths—AI for surface-level error correction and human-generated WCF for higher-order concerns such as structure, coherence, and argumentative depth (Bai & Nordin, 2025; Tran, 2025). Theoretically, such integration extends the WCF theory, socio-cognitive developmental models, and self-regulated learning models by investigating how different WCF modes scaffold learner growth. Moreover, with AI tools like ChatGPT and Gemini becoming increasingly prevalent in educational contexts (Tran, 2025), examining their effective integration is critical for future-oriented EFL pedagogy. Therefore, this study aimed to compare the effects of the human-generated, AI-generated, and combined WCF on Iranian EFL learners' writing performance, while exploring whether sequencing WCF (AI first, followed by human) improves writing quality and learner engagement compared to AI-

generated or human- generated WCF. To that end, this study addressed the following research questions:

> **RQ1:** Which WCF mode (i.e., AI-generated, teacher-generated, and combined WCF) produces greater improvement in intermediate-level Iranian EFL students' writing quality?
>
> **RQ2:** How do EFL learners perceive of AI-generated, human-generated, and combined WCF?

2. Literature Review

2.1. Writing in EFL Contexts

In EFL contexts, writing presents major challenges, particularly in grammar, organization, and coherence during academic writing. In countries such as Iran, learners often rely on formulaic expressions and surface understanding of grammar, while struggling with task-specific cohesion and content organization (Asadi et al., 2025; Kamali et al., 2024; Marzuki et al., 2023; Mohammadkarimi & Qadir, 2025). These challenges highlight the need for effective WCF mechanisms that address not only linguistic accuracy but also higher-order writing skills (Polakova & Ivenz, 2024; Steiss et al., 2024).

WCF plays a central role in learners' writing development (Mohammed & Khalid, 2025). Whether teacher-provided or automated, it supports awareness and revision by targeting both surface-level issues and higher-order aspects of writing (Abduljawad, 2025; Ding & Zou, 2024). Automated writing evaluation (AWE) tools (e.g., Grammarly, Pigai, and Criterion) deliver prompt feedback on surface-level features (e.g., syntax, spelling, and etc.), showing consistent efficacy in improving these areas (Escalante et al., 2023; Fleckenstein et al., 2023; Rahimi et al., 2025; Shi & Aryadoust, 2022). However, such tools cannot adequately address global concerns such as rhetorical structure, coherence, and creative development, which require nuanced, context-sensitive human insight (Ding & Zou, 2024).

Teacher-generated WCF, especially when dialogic, remains essential as it provides tailored guidance on textual structure, rhetorical strategy, and argumentative depth, which extend beyond the capacity of AI tools (Abduljawad, 2025). Although time-intensive, particularly in large EFL classrooms, teacher input enables adaptive scaffolding responsive to learners' immediate needs (Mahapatra, 2024). Peer feedback and self-assessment also encourage learner autonomy and engagement; however, they often lack the depth and consistency of expertise offered by trained instructors (Ding & Zou, 2024).

Generative AI systems, including Grammarly, ChatGPT, and other tools, deliver instant, scalable WCF on grammar, mechanics, and vocabulary. Some meta-analyses have confirmed their positive effects on learner attitudes and linguistic accuracy (Guan et al., 2024; Shi & Aryadoust, 2022). ChatGPT specifically has been applied in academic writing to provide interactive dialogue-based CF, fostering scaffolding and revision (Ding & Zou, 2024; Xu, 2025; Werdiningsih et al., 2024). Learners have reported advantages such as improved vocabulary, greater autonomy, and reduced linguistic anxiety, though concerns remain regarding originality and contextual sensitivity (Abduljawad, 2025).

The integration of WCF modes draws on socio-cultural, cognitive, and noticing-based theories. WCF acts as a zone of proximal development (ZPD) scaffold. Human instructors deliver tailored scaffolding for learners' needs, while AI provides frequent and routine prompts for self-regulatory development (Abduljawad, 2025). Effective WCF is dialogic and learner-centered. Whereas AI typically delivers surface-level corrections, teachers contribute interpretive, content-based, and evaluative feedback (Ding & Zou, 2024; Han et al., 2023a, 2023b). Writing development improves when learners consciously notice errors. The AI-generated WCF immediately draws attention to linguistic forms, though it may overlook discourse-level noticing (Yoon et al., 2023). Human-generated WCF, in contrast, extends noticing beyond error correction to global issues.

Together, these theories situate AI- and human-generated WCF as complementary rather than competitive: AI facilitates noticing and correction, while human input scaffolds coherence, voice, and argumentation. Dialogue-based and comparative studies imply that optimal writing development emerges when both instructors and AI function within learners' ZPD.

2.2. Empirical Studies

Research comparing the AI- and human-generated WCF emphasizes both strengths and weaknesses. Abduljawad (2025), for instance, compared ChatGPT with traditional instructor-generated WCF in Saudi EFL learners: while AI excelled in grammar and vocabulary, instructors better supported creativity, coherence, and contextual appropriateness. Likewise, Bagheri Nevisi and Mohammadi (2024) investigated the impact of automated writing evaluation (Grammarly) on Iranian EFL learners' essays. Findings showed significant improvement, reduced errors, increased enthusiasm, and preference for combining AWE with teacher-generated WCF, highlighting implications for pedagogy, curriculum design, and material development.

The meta-analysis by Ding and Zou (2024) revealed that AWE tools surpassed humans in addressing surface errors but were less effective for cohesion and argumentation. Similar patterns emerge in dialogue-based studies: Han et al. (2023) and Yoon et al. (2023) demonstrated how ChatGPT aids revision through conversational scaffolding, yet its WCF often lacks depth and remains generic. Collectively, these comparative insights suggest that AI support is effective for mechanical accuracy, while instructor-generated WCF is indispensable for higher-order aspects of writing, reinforcing the potential for integrative use, which is still underexplored.

Although evidence is growing, very few studies have adopted integrated sequences combining AI and human input in a single intervention. Much research remains divided, focusing either on the AI-generated WCF (e.g., Ding & Zou, 2024; Yoon et al., 2023) or human-generated WCF (Abduljawad, 2025). Mixed-method datasets such as RECIPE and ChEDDAR (Han et al., 2023) show promise but lack experimental rigor needed for comparing integrated versus isolated sequences. Few studies have systematically compared WCF sequencing (AI-first vs. human-first). This study addresses this gap by examining the AI-before-human sequence and its effects on writing performance and learner engagement. This review highlights that AI- and human-generated WCF types occupy overlapping but distinct areas of writing development. However, research has not yet explored collaborative pedagogical models that integrate both strengths. This study sought to address this gap by investigating integrated WCF sequences that address both higher-order skills and linguistic accuracy within the Iranian EFL context.

3. Method

3.1. Design

This study employed a mixed-methods quasi-experimental design to examine the impact of the AI-generated, human-generated, and mixed (AI + human) WCF on EFL writing. The pretest and posttest scores were compared across three groups. Because of curriculum limitations, a control group could not be included; however, the design allowed for both within-group and between-group comparisons. This design aligns with recent EFL research using ChatGPT (e.g., Mahapatra, 2024; Ekizoğlu & Demir, 2025, Zou et al., 2024). In addition, focus groups were used to capture learner experiences, following the ChEDDAR and RECIPE models.

3.2. Participants

Two participant groups were involved: one for quantitative analysis and one for qualitative interviews. For the quantitative phase, a total of 384

intermediate EFL learners (ages 18–24, ~52% female) were randomly sampled from two Iranian universities using Cochran's formula. The Oxford Quick Placement Test (scores: 60–70%) was used to verify competence. Participants were then assigned to three equivalent groups (n = 128): human-generated, AI-generated, and combined WCF. For the qualitative phase: 30 participants (10 per group) were purposively sampled by gender, participation, and writing performance to maximize diversity. Interviews continued until data saturation was reached (≈25 participants).

To control confounding variables, those with advanced writing experience or prior use of AI tools (e.g., ChatGPT, Grammarly) were excluded. Background questionnaires collected demographic data and learning styles, while randomization balanced group characteristics.

3.3. Instruments

Four instruments were used to measure the effects of WCF types: a) standardized rubric writing assignments; b) WCF generated by AI via ChatGPT; c) human-generated WCF using a protocol; and d) post-intervention focus group interviews.

3.3.1. Standardized Rubric Writing Assignments

The participants were asked to write argumentative essays (300–350 words) on academic topics (e.g., social media and learning). One essay was written before the intervention and one after. Topics were culturally neutral and cognitively stimulating. Essays were scored on an IELTS-adapted rubric covering: a) task response; b) coherence and cohesion; c) lexical resource; and d) grammatical range and accuracy. Two trained raters scored independently. If scores differed by ≥1 band, a third rater resolved disagreements.

3.3.2. ChatGPT-Generated Written Corrective Feedback

Students in the AI and Combined groups received automated weekly WCF from ChatGPT (GPT-4). Prompts instructed the AI to review grammar, vocabulary, organization, and coherence (e.g., Please review this essay for grammar and logic. Use bullet points.). Original and revised drafts were submitted weekly, and all WCF sessions were monitored for correction patterns (e.g., grammar, cohesion). This procedure followed best practices from ChEDDAR and RECIPE.

3.3.3. A Protocol for Human-Generated Written Corrective Feedback

Four experienced EFL instructors provided comments for the human-generated and Combined WCF groups. Using an IELTS-based template,

teachers commented on: a) content and relevance; b) organization and cohesion; c) grammar and syntax; and d) lexical range. The WCF consisted of margin notes and final comments (~100–150 words per draft). Calibration workshops ensured rater consistency. Comments were returned within 3 days to support revision.

3.3.4. Focus Group Interviews

Six semi-structured focus groups (5 participants each; N=30) were conducted after the intervention to explore attitudes, revision strategies, and affective responses. Sample questions included:
- "How useful was the WCF?"
- "Did you revise differently based on the type of WCF?"
- "Did the WCF affect your confidence?"

Interviews were held in Persian, translated, audio-recorded, transcribed, and thematically analyzed. Themes were then compared with writing score patterns.

3.4. Procedure

The 10-week study unfolded in three phases: diagnostic (Week 1), intervention (Weeks 2–7), and post-assessment with qualitative follow-up (Weeks 8–10).

Week 1: All 384 learners wrote a pretest argumentative essay (300–350 words) in class under timed conditions (45 minutes, handwritten, no dictionaries or computers). This established baseline measures for task response, cohesion, vocabulary, and grammar.

Weeks 2–7 (Intervention Phase): Weekly assignments on academic topics were completed, with treatment differing by group:

- *AI-generated WCF group:* Drafts submitted to ChatGPT (GPT-4) with standardized prompts. Revisions were based on the AI-generated WCF. All logs were collected for analysis.
- *Human-generated WCF group:* Drafts reviewed by teachers, who provided margin notes and ~100–150-word summary WCF using a rubric. Students revised and resubmitted.
- *Combined WCF group:* Drafts were first revised based on AI-generated WCF, then submitted for teacher-generated WCF, which worked as sequential combined WCF.

Week 8: All students completed a posttest essay under identical conditions as the pretest for comparability.

Weeks 9–10: Six focus group interviews (N=30) explored experiences across WCF types, covering revision strategies, motivation, and perceptions.

The consistency measures were as follows. Standardized prompts were used with AI. Teachers applied the same rubric across groups. WCF was returned within 3 working days. Students submitted both drafts and revisions weekly. All essays, AI logs, and teacher-generated WCF were collected for analysis. This systematic procedure ensured consistency, internal validity, and comparability across groups.

Informed consent was obtained, confidentiality was assured, and pseudonyms replaced identifiers. Focus group recordings were made only with explicit permission. Ethical practices ensured transparency, voluntariness, and compliance with research standards.

3.5. Data Analysis

Writing ability was assessed using the IELTS rubric. Composite band scores were assigned for pretest and posttest essays. Using SPSS (v25), paired-samples t-tests measured within-group change. One-way ANOVA compared gain scores among groups. Tukey's HSD was used to identify significant between-group differences. Effect sizes were reported using Eta squared (ANOVA) and Cohen's d (t-tests), consistent with EFL studies (Ding & Zou, 2024; Ekizoğlu & Demir, 2025).

Six focus-group recordings were transcribed and coded in NVivo 14 using thematic analysis. Initial codes, informed by prior research (Han et al., 2023; Mahapatra, 2024), included clarity, motivation, WCF utility, and affective responses, while new themes such as AI fairness and trust in teachers emerged inductively. Two coders reached >85% agreement. Thematic synthesis examined how WCF influenced engagement, emotions, and revision strategies, with attention to student perceptions of AI tools.

To enhance validity, both quantitative and qualitative findings were compared. For example, AI group participants often described ChatGPT as "fast" and "clear," though sometimes "mechanical." The human-generated WCF group valued WCF as "personal" and "emotionally supportive," even when improvements were less dramatic. NVivo matrices mapped convergence and divergence, following methods from ChEDDAR and RECIPE studies to examine the alignment between quantitative and qualitative findings.

4. Results

This section presents the quantitative findings on the effect of the AI-generated, human-generated, and combined WCF on students' EFL writing performance. The pretest and posttest data were analyzed to examine within-group improvements and between-group differences in terms of writing gains.

Descriptive statistics, paired-samples t-tests, and one-way ANOVA were run to evaluate the statistical significance of the observed patterns.

4.1. Results for the First Research Question

A total of 384 participants were evenly distributed across three groups: AI-generated, human-generated, and combined WCF (AI- + human-generated). Each participant completed both a pretest and posttest writing task. Table 1 displays the mean scores and standard deviations for the pretest, posttest, and gain scores in the three groups.

Table 1
Descriptive Statistics for Pretest, Posttest, and Gain Scores by Group

Group	Pretest M(SD)	Posttest M(SD)	Gain M(SD)	t	p	Cohen's d	Effect Size Interpretation
AI	5.46(0.57)	6.09(0.55)	0.63(0.35)	17.67	< .001	1.52	Large
Human	5.62(0.59)	6.11(0.54)	0.49(0.38)	13.15	< .001	1.17	Medium to Large
Combined	5.47(0.59)	6.36(0.55)	0.89(0.36)	24.05	< .001	2.13	Very Large

As shown in Table 1, the AI-generated group's mean pretest score was 5.46, while the mean posttest score was 6.09, yielding a gain of approximately 0.63 points. This difference was statistically significant, $t(127) = 17.67$, $p < .001$. For the human-generated WCF group, the mean pretest score was 5.62, and the mean posttest score was 6.11, with a gain of about 0.49 points. This improvement was also statistically significant, $t(127) = 13.15$, $p < .001$. moreover, for the combined WCF group, the mean pretest score was 5.47, while the posttest score rose to 6.36, showing a gain of 0.89 points. The paired-samples t-test confirmed a significant difference, $t(127) = 24.05$, $p < .001$.

These results suggest that all WCF types led to statistically significant gains in their writing scores, with the combined WCF condition producing the highest improvement. In addition to statistical significance, effect sizes were calculated using Cohen's d to measure the magnitude of improvement. The AI-generated WCF group showed a large effect size ($d = 1.52$), the human-generated WCF group a medium effect size ($d = 1.17$), and the combined WCF group demonstrated a very large effect size ($d = 2.13$). This indicates substantial learning benefits across all groups, with the combined WCF group showing the most notable improvement. To compare the effectiveness of the three WCF conditions, a one-way ANOVA was conducted on the gain scores (Table 2).

Table 2

One-Way ANOVA on Gain Scores

Source of Variation	SS	df	MS	F	p	η^2
Between Groups	17.82	2	8.91	35.72	< .001	0.16
Within Groups	378.88	381	0.994			
Total	396.70	383				

Note: Effect size $\eta^2 = 0.16$ reflects a large between-group difference in writing improvement.

The results in Table 2 showed a statistically significant difference among the groups, $F(2, 381) = 35.72$, $p < .001$. Post-hoc analysis with Tukey's HSD test was also run to see where the differences lie (Table 3).

Table 3

Post-Hoc Comparisons of Gain Scores by WCF Condition (Tukey's HSD)

Comparison	Mean Difference (MD)	SE	P	95% CI [Lower, Upper]
Combined – AI-generated	0.26	0.05	< .001	0.16, 0.36
Combined – Human-generated	0.40	0.05	< .001	0.30, 0.50
AI-generated – Human-generated	0.14	0.05	< .05	0.04, 0.24

As revealed in Table 3, the combined WCF group outperformed both AI-generated and human-generated WCF groups significantly ($p < .001$). The AI-generated group achieved slightly higher gains than the human-generated group, and this difference was also statistically significant ($p < .05$). Thus, while all WCF types were effective, the combined WCF (AI-generated + human-generated) produced the largest improvement, followed by the AI-generated WCF, and then the human-generated WCF.

Overall, the effect size for the ANOVA, calculated using eta squared indicated a large effect of WCF type on writing performance gains, according to Cohen's (1988) conventional benchmarks ($\eta^2 = 0.16$). This further emphasizes the substantial role of WCF modality in improving learners' writing proficiency. A bar chart shown in Figure 1 illustrates the mean gain scores across the three WCF groups.

Figure 1
Mean Gain Scores by WCF Group
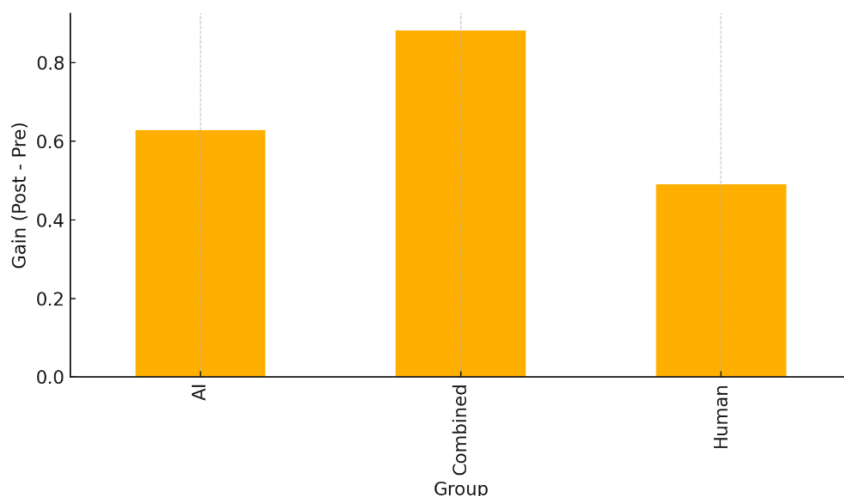
As presented in Figure 1, the combined WCF group showed the highest gain (approximately 0.89 points), compared to 0.63 for the AI-generated WCF group, and 0.49 for the human-generated WCF group. This visual representation highlights the relative advantage of a hybrid WCF approach in enhancing EFL learners' writing proficiency and supports the statistical findings. Drawn from both descriptive and inferential statistics, these findings provide strong evidence that combining the AI-generated and teacher-generated WCF yields superior outcomes in writing development.

4.2. The Results for the Second Research Question

To complement the quantitative findings, the thematic analysis of focus group data revealed four major themes and seven subthemes reflecting learners' perceptions of the AI-generated, human-generated, and combined WCF. These themes provide insights into how students engaged with, processed, and emotionally responded to the different WCF modalities.

Participant A (AI-generated WCF group): "ChatGPT helped me fix grammar mistakes very fast. It was good because I didn't feel embarrassed asking again."

Participant B (Combined WCF group): "First, I used the AI to clean the grammar, then I gave it to the teacher. She focused on ideas and structure. That helped me the most."

Participant C (Human-generated WCF group): "My teacher always explained why something was wrong, which I liked. But sometimes the comments were too long and I didn't know where to start."

Participant D (AI-generated WCF group): "The robot was fast, but sometimes I didn't understand what it meant by 'awkward sentence'."

These quotes illustrate the nuanced ways students internalized and acted upon different forms of WCF. Notably, those in the combined WCF group consistently described a clearer editing workflow and stronger confidence in their writing revisions. The analysis of the interviews identified 128 coded references across all interviews. The most frequent codes were: a) clarity of the AI-generated WCF (28 references), b) teacher explanation helpful (23 references), c) AI fast but generic (19 references), d) combination best (16 references), and e) AI reduces anxiety (14 references). The sample codes are shown in Table 4.

Table 4
Sample Codes, Themes, and Representative Quotes from Focus Group Analysis

Theme	Code example	Participant Quote
WCF clarity	clear structure	"ChatGPT helped me fix grammar mistakes very fast." (AI-generated)
Perceived fairness	robot doesn't judge	"AI was fair; it didn't criticize me like a person." (Combined)
Motivation and anxiety	less fear of judgment	"I felt less nervous with AI—it doesn't judge." (AI-generated)
Revision strategies	AI first, then refine	"I used AI first, then my teacher helped improve my ideas." (Combined)

These patterns support the quantitative findings, showing that while AI tools support rapid revision, human-generated WCF provides conceptual depth. The combination of both WCF types appears to maximize clarity, confidence, and revision quality. Overall, the findings suggest that while AI supports efficiency and affective benefits, human-generated WCF provides essential scaffolding, and combining both maximizes learners' writing development.

5. Discussion

This mixed methods study investigated the comparative effects of the AI-generated, human-generated, and combined (AI + human) WCF on the academic writing performance of intermediate Iranian EFL learners. The results revealed statistically significant improvements in all three groups, as

confirmed by paired-samples t-tests; however, the extent of improvement varied across conditions. The combined WCF group demonstrated the most substantial gain, followed by the AI-generated group, and the human-generated group.

Effect sizes further highlighted these differences: combined WCF yielded a very large effect, the AI-generated a large effect, and the human-generated a moderate-to-large effect. One-way ANOVA on gain scores indicated a significant difference among the three groups with a large effect size, confirming that WCF modality substantially impacted learning outcomes. These findings suggest that while each WCF type supports learner progress, the synergistic use of the AI-generated and teacher-generated WCF leads to deeper and more comprehensive writing development.

These results align with the noticing hypothesis (Schmidt, 1990), as AI provides immediate and clear error highlighting that enhances learner awareness, whereas the teacher-generated WCF supports scaffolded learning within Vygotsky's (1978) ZPD, particularly in developing coherence, argumentation, and genre awareness. The study's findings also resonate with prior research such as Abduljawad (2025), who observed that the AI-generated WCF is particularly effective for grammatical and lexical issues, consistent with our results where the AI-assisted students improved rapidly in surface-level linguistic accuracy. In contrast, human-generated WCF was more effective in structuring ideas, refining coherence, and supporting argumentative clarity, a pattern observed in our study despite comparatively lower overall gains.

The superiority of the combined WCF condition aligns with Mahapatra (2024), who reported that layered WCF—AI suggestions followed by teacher input—maximized learning gains. Similarly, Ekizoğlu and Demir (2025) found that dual-modality WCF allowed learners to balance micro-level corrections (e.g., grammar, vocabulary) with macro-level revisions (e.g., structure, thesis development). Our study extends these findings by quantitatively confirming that hybrid WCF produces significantly greater effect sizes than either approach alone, providing empirical support for WCF (Ellis, 2009), which emphasizes that WCF is most effective when timely, clear, and appropriately scaffolded—qualities achieved when AI and human input are strategically sequenced.

Moreover, the findings enrich the noticing hypothesis framework by showing that the AI-generated WCF's immediacy likely enhanced learners' awareness of form, while human-generated WCF encouraged deeper cognitive engagement through elaboration and dialogic explanation, as reflected in focus group responses. Qualitative data indicated that learners in the combined WCF

group experienced reduced anxiety and increased confidence, particularly when the AI-generated WCF preceded teacher intervention. These observations validate the quantitative superiority of this group: the real-time AI-generated WCF lowered affective barriers, making learners more receptive to subsequent human guidance. Students receiving only human-generated WCF often noted delays or imprecision in teacher responses, which may explain their comparatively lower gains. Thematic analysis also suggested that students benefiting from the AI-generated WCF reported increased autonomy and openness, which could account for their strong performance on surface-level measures such as vocabulary and grammar. This convergence of qualitative and quantitative evidence underscores the interplay of cognitive and affective variables in mediating WCF efficacy.

Additionally, the study provides indirect validation for Vygotsky's (1978) sociocultural theory, emphasizing the role of interaction and mediation in cognitive development. Human-generated WCF functions as mediational scaffolding supporting self-regulation, whereas AI tools serve as cognitive artifacts extending learners' ZPDs during the writing process. The combined modality reflects a blended mediation framework, where technology and human instruction operate together to foster development (Godwin-Jones, 2025). This study implemented only AI-first followed by human-generated WCF; future research should explore alternative sequences to examine whether order affects WCF uptake and writing growth.

While these theoretical frameworks sometimes diverge—e.g., the Noticing Hypothesis centers on individual cognitive awareness, whereas Sociocultural Theory emphasizes dialogic negotiation—they can be reconciled through the strategic integration of the AI- and human-generated WCF. AI enhances noticing and salience, while the teacher-generated WCF elaborates, scaffolds, and dialogizes learning. The degree to which these frameworks explain learner development also depends on proficiency level, task difficulty, and learner autonomy. An integrated framework combining cognitive, sociocultural, and instructional WCF paradigms offers the most explanatory power for writing development in AI-enhanced contexts.

Some prior studies report mixed results for AWE tools used in isolation. Ding and Zou (2024) noted that AI systems like Grammarly improve grammatical accuracy but often miss contextual or rhetorical appropriateness, leading to surface-level gains. The lower effect size of the AI-generated WCF in our study supports this critique. Nevertheless, AI-generated WCF alone, when guided through pedagogically designed prompts (e.g., GPT-4), can yield substantial improvement—a result not consistently observed in earlier rule-

based AWE systems. Overall, this study contributes to the growing argument that the AI-generated WCF is not merely supplemental but can be a dynamic learning tool, especially when embedded within human-supervised frameworks that consider learner agency, WCF uptake, and instructional alignment.

6. Conclusions and Implications

This study examined the comparative effect of the AI-generated, human-generated, and combined WCF on the writing performance of intermediate Iranian EFL learners using a mixed-methods quasi-experimental design. All groups showed significant improvement, with combined WCF yielding the highest gains, followed by the AI-generated and human-generated WCF. In conclusion, combining the AI-generated and human-generated WCF produces higher EFL writing gains compared to either modality. Linking these outcomes to theoretical frameworks and practical recommendations provides a robust foundation for pedagogical innovation in the AI-enhanced language learning.

Key findings suggest AI-generated WCF is highly effective for surface-level features (grammar, vocabulary) due to its immediacy and clarity, while human-generated WCF more effectively supports idea development, organization, and coherence, which is consistent with sociocultural and cognitive learning theories. When combined, AI-generated and human-generated WCF produce synergistic effects, enhancing both linguistic accuracy and rhetorical quality.

The implications for language educators and policymakers include designing hybrid WCF structures by integrating AI tools (e.g., ChatGPT, Grammarly) as initial aids, allowing teachers to focus on global commentary and higher-order concerns. Professional teacher development is also essential to train the AI-generated corrections and target argumentation, style, and content depth, thereby complementing automated WCF. Curricular reforms should implement AI-enhanced writing tracks, where automated WCF is followed by teacher-mediated refinement to promote autonomy and skill progression. Additionally, the efficient use of instructional time can be achieved as AI manages low-level errors, freeing class time for peer review, workshops, and seminar-style discourse to enrich engagement.

The study was limited to one language proficiency level; therefore, the findings may differ for novice or advanced learners. The selected AI tool may not reflect future versions, and instructor WCF is subject to variability in expertise and judgment. Despite randomization and participant selection, statistical controls (e.g., ANCOVA) were not employed to isolate potential confounding variables such as prior digital literacy, learning styles, or attitudes

toward WCF. Differences in motivation or technology comfort, previous AI exposure, and management of WCF were not controlled, limiting the strength of statistical inferences in this study. Future studies should control these variables as covariates. Cultural and educational backgrounds also influence WCF reception and utilization. Future research should include mixed proficiency levels and sociocultural contexts to improve generalizability. Additionally, future studies could explore WCF sequencing and timing, interactions with learner proficiency, longitudinal efficacy, deeper qualitative insights via classroom observation or case studies, and hybrid WCF models across multilingual and multicultural settings.

References

Abduljawad, S. A. (2025). Investigating the impact of ChatGPT as an AI tool on ESL writing: Prospects and challenges in Saudi Arabian higher education. *International Journal of Computer-Assisted Language Learning and Teaching, 14*(1), 1–19. https://doi.org/10.4018/IJCALLT.367276

Asadi, M., Ebadi, S., & Mohammadi, L. (2025). The impact of integrating ChatGPT with teachers' feedback on EFL writing skills. *Thinking Skills and Creativity, 101766.* https://doi.org/10.1016/j.tsc.2025.101766

Bai, X., & Nordin, N. R. M. (2025). Human-AI collaborative feedback in improving EFL writing performance: An analysis based on natural language processing technology. *Eurasian Journal of Applied Linguistics, 11*(1), 1–19. https://doi.org/10.32601/ejal.11101

Bagheri Nevisi, R., & Mohammadi, R. (2024). The impact of automated writing evaluation on Iranian EFL learners' essay writing: A mixed-methods study. *Mixed-Methods Studies in English Language Teaching, 1*(1), 23–46. https://doi.org/10.71873/mslt.2024.1121001

Cao, S., & Zhong, L. (2023). Exploring the effectiveness of ChatGPT-based feedback compared with teacher feedback and self-feedback: Evidence from Chinese to English translation [Preprint]. *arXiv.* https://doi.org/10.48550/arXiv.2309.01645

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge.

Ding, L., & Zou, D. (2024). Automated writing evaluation systems: A systematic review of Grammarly, Pigai, and Criterion with a perspective on future directions in the age of generative artificial intelligence. *Education and Information Technologies, 29*(11), 14151–14203. https://doi.org/10.1007/s10639-023-12402-3

Ekizoğlu, M., & Demir, A. (2025). AI-assisted writing feedback for enhancing secondary students' writing skills: An experimental study [Preprint]. *Research Square.* https://doi.org/10.21203/rs.3.rs-6430737/v1

Escalante, J., Pack, A., & Barrett, A. (2023). AI-generated feedback on writing: Insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education, 20,* 57. https://doi.org/10.1186/s41239-023-00425-2

Fleckenstein, J., Liebenow, L. W., & Meyer, J. (2023). Automated feedback and writing: A multi-level meta-analysis of effects on students' performance. *Frontiers in Artificial Intelligence, 6,* 1162454. https://doi.org/10.3389/frai.2023.1162454

Godwin-Jones, R. (2025). *Towards sustainable technology use in instructed second language acquisition.* John Benjamins.

Guan, L., Li, S., & Gu, M. M. (2024). AI in informal digital English learning: A meta-analysis of its effectiveness on proficiency, motivation, and self-regulation. *Computers and Education: Artificial Intelligence, 7,* 100323.

Han, J., Yoo, H., Myung, J., Kim, M., Lee, T., Ahn, S.-Y., & Oh, A. (2023a). ChEDDAR: Student–ChatGPT dialogue in EFL writing education [Preprint]. *arXiv.* https://doi.org/10.48550/arXiv.2309.13243

Han, J., Yoo, H., Myung, J., Kim, M., Lim, H., Kim, Y., Lee, T., Hong, H., Kim, J., & Oh, A. (2023b). LLM as a tutor in EFL writing education: Focusing on the evaluation of student–LLM interaction [Preprint]. *arXiv.* https://doi.org/10.48550/arXiv.2310.05191

Kamali, J., Paknejad, A., & Poorghorban, A. (2024). Exploring the challenges and affordances of integrating ChatGPT into language classrooms from teachers' points of view: An ecological perspective. *Journal of Applied Learning and Teaching, 7*(2). https://doi.org/10.37074/jalt.2024.7.2.8

Karagöz, I. (2025). AI-generated feedback in English writing instruction for language learners: A systematic review. *The Reading Matrix, 25*(1), 51–67.

Lee, S., Choe, H., Zou, D., & Jeon, J. (2025). Generative AI (GenAI) in the language classroom: A systematic review. *Interactive Learning Environments,* 1–25. https://doi.org/10.1080/10494820.2025.2498537

Li, J., Huang, J., Wu, W., & Whipple, P. B. (2024). Evaluating the role of ChatGPT in enhancing EFL writing assessments in classroom settings: A preliminary investigation. *Humanities and Social Sciences Communications, 11,* Article 1268. https://doi.org/10.1057/s41599-024-03755-2

Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: A mixed methods intervention study. *Smart Learning Environments, 11,* Article 9. https://doi.org/10.1186/s40561-024-00295-9

Marzuki, W., Widiati, U., Rusdin, D., Darwin, & Indrawati, I. (2023). The impact of AI writing tools on the content and organization of students' writing: EFL teachers' perspective. *Cogent Education, 10*(2), 2236469. https://doi.org/10.1080/2331186X.2023.2236469

Mohammadkarimi, E., & Qadir, B. M. (2025). The impact of artificial intelligence use on students' autonomous writing. *Journal of Applied Learning and Teaching, 8*(1), 143–153. https://doi.org/10.37074/jalt.2025.8.1.14

Mohammed, S. J., & Khalid, M. W. (2025). Under the world of AI-generated feedback on writing: Mirroring motivation, foreign language peace of

mind, trait emotional intelligence, and writing development. *Language Testing in Asia, 15*(7). https://doi.org/10.1186/s40468-025-00343-2

Polakova, P., & Ivenz, P. (2024). The impact of ChatGPT feedback on the development of EFL students' writing skills. *Cogent Education, 11*(1), 2410101. https://doi.org/10.1080/2331186X.2024.2410101

Pratama, A., & Sulistiyo, U. (2024). A systematic review of artificial intelligence in enhancing English foreign learners' writing skills. *International Journal of Education, 3*(2), 170–181. https://doi.org/10.2829/ipej.2024.320170

Rahimi, M., Fathi, J., & Zou, D. (2025). Exploring the impact of automated written corrective feedback on the academic writing skills of EFL learners: An activity theory perspective. *Education and Information Technologies, 30,* 2691–2735. https://doi.org/10.1007/s10639-024-12896-5

Shi, H., & Aryadoust, V. (2024). A systematic review of AI-based automated written feedback research. *ReCALL, 36*(2), 187–209. https://doi.org/10.1017/S0958344023000265

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*(2), 129–158. https://doi.org/10.1093/applin/11.2.129

Steiss, J., Tate, T., Graham, S., Cruz, J., Hebert, M., Wang, J., Moon, Y., Tseng, W., Warschauer, M., & Olson, C. B. (2024). Comparing the quality of human and ChatGPT feedback on students' writing. *Learning and Instruction, 91,* Article 101894. https://doi.org/10.1016/

Tran, T. T. T. (2025). Enhancing EFL writing revision practices: The impact of AI- and teacher-generated feedback and their sequences. *Education Sciences, 15*(2), Article 232. https://doi.org/10.3390/educsci15020232

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

Wang, D. (2024). Teacher versus AI-generated corrective feedback and language learners' writing anxiety, complexity, fluency, and accuracy. *International Review of Research in Open and Distributed Learning, 25*(3), 37–56. https://doi.org/10.19173/irrodl.v25i3.7646

Werdiningsih, I., Marzuki, W., & Rusdin, D. (2024). Balancing AI and authenticity: EFL students' experiences with ChatGPT in academic writing. *Cogent Arts & Humanities, 11*(1). https://doi.org/10.1080/23311983.2024.2392388

Xu, Z. (2025). Patterns and purposes: A cross-journal analysis of AI tool usage in academic writing [Preprint]. *arXiv.* https://doi.org/10.48550/arXiv.2502.00632

Yoon, S.-Y., Miszoglad, E., & Pierce, L. R. (2023). Evaluation of ChatGPT feedback on ELL writers' coherence and cohesion [Preprint]. *arXiv.* https://doi.org/10.48550/arXiv.2310.06505

Zou, S., Guo, K., Wang, J., & Liu, Y. (2024). Investigating students' uptake of teacher- and ChatGPT-generated feedback in EFL writing: A comparison study. *Computer Assisted Language Learning,* 1–30. https://doi.org/10.1080/09588221.2024.2447279