Language Teaching Research Quarterly **EUROKD**

2026 (54), 102–124
https://doi.org/10.32038/ltrq.2026.54.05

Optimizing EFL Article Accuracy with Hybrid AI-Teacher Feedback

Afshin Soori
Department of English Language and Literature, Larestan Branch, Islamic Azad University, Larestan, Iran
Laleh Khojasteh*
Department of English Language, School of Paramedical Sciences, Shiraz University of Medical Sciences, Shiraz, Iran
Sedigheh Shakib Kotamjani
Department of English Language Teaching and Educational Management, Central Asian University, Tashkent, Uzbekistan
Sanaz Mojarrad
Department of Persian Language and Literature, Shiraz University of Medical Sciences, Shiraz, Iran

Correspondence
Email: khojastehlaleh@yahoo.com

Abstract

This study examines the effectiveness of various feedback models in enhancing grammatical accuracy in English article usage among 140 Iranian EFL medical students. Using a randomized experimental design, participants were assigned to five groups: AI feedback alone (n=28), AI with immediate oral teacher feedback (n=30), AI with delayed written teacher feedback (n=35), a hybrid model combining AI with both teacher feedback types (n=24), and a control group receiving traditional feedback (n=23). To evaluate performance, we administered pre-tests, post-tests, and delayed post-tests, while a post-study survey gauged learner attitudes. The data revealed that the hybrid model yielded the most substantial improvements in accuracy and long-term retention (post-test M=21.83; delayed post-test M=21.29). The next most effective condition was AI coupled with immediate oral feedback. Notably, AI feedback used in isolation surpassed AI with delayed written feedback, underscoring the crucial role of immediacy in learning. Furthermore, the survey showed a strong positive correlation between learner attitudes and performance ($r=.32$, $p<.001$), with the hybrid group expressing the most positive perceptions. These findings highlight the pedagogical value of integrating AI tools with timely, multi-modal teacher feedback to foster linguistic development and learner engagement in EFL contexts.

ARTICLE HISTORY
Received: 30 May 2025
Revised: 22 February 2026
Accepted: 15 March 2026

KEYWORDS
AI Feedback, EFL learners, Hybrid Feedback, Grammatical Accuracy, Learner Attitudes

How to cite this article (APA 7th Edition):
Soori, A., Khojasteh, L., Kotamjani, S. S., & Mojarrad, S. (2026). Optimizing EFL article accuracy with hybrid AI-teacher feedback. *Language Teaching Research Quarterly*, *54*, 102–124. https://doi.org/10.32038/ltrq.2026.54.05

Introduction

Feedback is a cornerstone of second language acquisition (SLA), enabling learners to identify and correct errors, thereby fostering accurate language skills (Karim & Nassaji, 2019). It plays a critical role in transforming input into intake, guiding learners toward more accurate language production through scaffolded instruction and metalinguistic reflection (Ellis, 2012; Lyster & Saito, 2010). Traditional teacher feedback, often timely and context-specific, has been shown to significantly enhance grammatical accuracy and overall proficiency, especially when delivered in a focused and interactive format (Bitchener & Storch, 2016; Han & Hyland, 2015).

However, the rapid advancement of digital technologies has introduced artificial intelligence (AI)-generated feedback as a complementary and increasingly accessible tool in SLA environments. AI offers immediacy, scalability, and consistency—attributes that are challenging to replicate in traditional feedback settings (Hwang & Chien, 2022; Jegede, 2024). These developments align with calls for more personalized and data-driven feedback systems in computer-assisted language learning (Naseer & Khawaja, 2025). Leveraging natural language processing and large-scale error databases, AI systems excel at identifying syntactic and lexical inaccuracies, often delivering corrective feedback in real-time (Liang et al., 2023; Shen & Chong, 2023). Recent research has moved beyond error correction to explore learners' cognitive and emotional engagement with AI-generated feedback. For example, Dong (2024) proposed a ChatGPT-based feedback engagement framework that conceptualizes feedback interaction as both a cognitive and affective process, emphasizing the need for reflection and trust in AI-mediated learning. Similarly, Crosthwaite and Sun (2026) reviewed L2 writing feedback studies involving generative AI and noted that most prior work focused narrowly on surface-level errors, calling for greater attention to grammar-specific outcomes and long-term retention—issues central to the present study.

Despite these advantages, AI-generated feedback is not without limitations. Its lack of affective interaction and socio-cognitive sensitivity can reduce opportunities for deep processing and learner reflection—factors known to influence long-term retention (Liang et al., 2023; Liu et al., 2023). Conversely, teacher feedback, though rich in contextual nuance and learner-specific adaptation, remains constrained by issues of time, workload, and consistency (Bitchener & Storch, 2016; Nassaji & Kartchava, 2020). These contrasting affordances have prompted growing interest in hybrid feedback systems, which integrate the strengths of both AI tools and human instruction (Otaki, 2023; Woodworth, 2022). Recent empirical work supports this convergence: Bai and Nordin (2025) found that a human–AI collaborative feedback model significantly improved EFL learners' writing accuracy, complexity, and fluency, with cognitive engagement and writing enjoyment mediating these effects. Likewise, Lo et al. (2025) demonstrated that combining teacher comments with AI feedback enhanced learners' perceptions of feedback credibility and improved revision quality. Such findings echo Banihashem et al. (2025), who proposed a learner-centered framework for hybrid intelligent feedback that integrates teacher mediation, learner agency, and iterative reflection. Further supporting the potential of hybrid approaches, a meta-analysis by Kaliisa et al. (2026) confirmed that AI-teacher integrated feedback systems consistently outperform either modality alone in grammar learning contexts, attributing this to the complementary nature of their strengths.

A growing body of research indicates that a synergy between AI-generated and teacher-provided feedback can yield greater improvements in grammatical accuracy than either approach in isolation, especially when feedback is delivered in a timely, scaffolded, and contextually relevant manner (Wei-Xun & Jia-Ying, 2024; Zheng et al., 2024). More recent studies (e.g., Dong, 2024; Crosthwaite & Sun, 2026) have further emphasized that effective hybrid feedback should not only correct language but also foster reflective engagement, a factor closely linked to sustained accuracy and learner autonomy. However, the existing literature has explored this synergy in contexts secondary to core grammatical instruction. For instance, Namaziandost and Rezai (2024) investigated the function of AI within open and distributed learning frameworks, while Wei-Xun and Jia-Ying (2024) work centered on vocabulary development through AI-powered applications. This leaves a noticeable gap in our understanding of how hybrid feedback models might address one of the most intransigent grammatical challenges for EFL learners: the correct use of articles. Given the infamous complexity and rule-defying nature of English articles (Tsao et al., 2021; Zarei et al., 2020), this area remains critically underexplored. By situating the current study within this emerging body of research, it directly responds to recent calls for grammar-focused investigations that evaluate not only the cognitive impact of hybrid feedback but also learners' affective responses to it. Notably, Siregar et al. (2026) specifically highlighted the lack of research on hybrid feedback for complex grammatical features like articles, reinforcing the need for targeted studies in this area.

Beyond the mechanics of feedback delivery, how learners perceive the feedback they receive plays a critical role in its ultimate success. Research consistently shows that when students hold positive attitudes toward feedback, they are more motivated, engaged, and more likely to improve their grammatical skills (Tsao et al., 2021; Van der Kleij & Lipnevich, 2021). This presents a potential challenge for AI-based tools, as students sometimes view automated feedback as impersonal or less trustworthy compared to guidance from a teacher (Denny et al., 2023), potentially limiting its impact, particularly in professional identity formation (Nasiri & Shokrpour, 2024). Therefore, integrating AI suggestions with direct teacher input could be a powerful solution. This blended approach may improve the standing of the AI tool in students' eyes, building the trust and motivation necessary for effective learning (Otaki, 2023; Zarei et al., 2020). Supporting this view, Huang and Mizumoto (2025) found that generative AI can effectively mediate the relationship between learners' motivational self-systems and their engagement with written feedback, highlighting the importance of teacher-supported interpretation to sustain trust and motivation in AI-assisted feedback settings. Recent work by Andriani et al. (2026) provides empirical evidence for this, demonstrating that teacher scaffolding of AI feedback significantly enhances learner trust and perceived usefulness, particularly for complex grammatical corrections, directly linking positive perceptions to improved article accuracy.

To address the previously identified gap in the literature, our study was designed to systematically evaluate a novel hybrid feedback model. This model combines the immediate corrective power of AI with the nuanced guidance of a teacher, delivered through both immediate oral and delayed written formats. We chose to focus this investigation on the acquisition of English articles, a well-known stumbling block for EFL learners that is nonetheless fundamental to achieving grammatical precision. The research design contrasts a control group with four distinct feedback conditions: (1) AI feedback alone, (2) AI with immediate oral teacher feedback, (3) AI with delayed written teacher feedback, and (4) a fully integrated hybrid model. Through this comparison, we aim to measure not only immediate learning gains but also the crucial factor of long-term retention. Finally, recognizing that learner buy-in is essential, we also analyze students' perceptions of each feedback method to understand how their attitudes relate to their performance.

Research Questions

The study addresses the following key research questions:
RQ_1: How effective is AI feedback, alone or combined with teacher feedback, in enhancing article usage?
RQ_2: Which combination yields the greatest accuracy and retention?
RQ_3: How do students' perceptions of feedback relate to their performance?

Methodology

Research Design and Participants

This study employed a randomized controlled experimental design to examine the effects of various feedback types on the grammatical accuracy of English article usage among EFL learners. The research was conducted during the spring semester of 2024 at Shiraz University of Medical Sciences, Iran. The participants were 140 first-year Iranian medical students enrolled in a compulsory three-credit academic writing course at Shiraz University of Medical Sciences. Students were at a B1–B2 proficiency level (based on the university's placement test, aligned with the CEFR framework) and had comparable exposure to English instruction in prior semesters, including Pre-University, General

English 1, and General English 2 courses. Instruction was delivered by two experienced EFL instructors with over ten years of teaching experience, who were trained to ensure uniform implementation of the feedback procedures. They were randomly assigned to five groups, including a control group that received traditional pen-and-paper corrective feedback. The other four experimental groups received different combinations of AI-generated, immediate oral, and delayed written teacher feedback. AI feedback was provided through Grammarly Premium (v. 2024.12), a natural language processing–based writing assistant, which offered automated corrective feedback on article use, word choice, and sentence structure. Teachers monitored and clarified AI-generated suggestions during class discussions to prevent misunderstanding or over-reliance on the tool. Group sizes varied slightly due to enrollment constraints but remained balanced across conditions (see Table 1). The experimental procedure spanned nine weeks and included pre-test, post-test, and delayed post-test assessments to measure immediate learning outcomes and long-term retention. All writing tasks were short argumentative or descriptive paragraphs (120–150 words) related to the course topics, completed both in class and via the university's platform, where AI feedback was accessible. The same task prompts were used across all groups to ensure comparability.

Table 1

Number of Students in Experimental and Control Groups

Groups		N
Ex1: AI feedback alone	Immediate AI-generated corrective feedback on writing tasks.	28
Ex2: AI Feedback + Immediate Oral Teacher Feedback	AI feedback is immediately followed by oral corrective feedback from the teacher.	30
Ex3: AI Feedback + Delayed Teacher Feedback (Written)	AI feedback followed by written teacher feedback after a 24-48 hour delay.	35
Ex4: AI Feedback + Immediate Oral Feedback + Delayed Written Feedback	AI feedback, immediate oral feedback, and written feedback after a delay.	24
Control	Traditional pen and paper corrective feedback	23
Total		140

This study involved human participants in a classroom setting. The research was reviewed by the Institutional Review Board (IRB) at Central Asian University and was approved under Ethical Review Code: CAU-IRB-2025-032, as it comprised standard EFL writing tasks with no physical, psychological, or privacy risks. All participants were informed of the study's purpose, and their responses were anonymized (test scores only). All collected data were anonymized and securely stored, ensuring no personal identities were disclosed, in alignment with the ethical principles outlined in the Declaration of Helsinki (2013).

Instruments

To gather the necessary data for this investigation, we utilized a set of three primary instruments: a series of rational cloze tests to measure grammatical accuracy, a set of writing tasks to elicit article usage in a productive context, and an exit questionnaire to capture learner perceptions.

Rational cloze tests

Participants' declarative knowledge of English articles was assessed using three rational cloze tests administered at pre-test, post-test, and delayed post-test stages. This instrument, valued for its high reliability in measuring grammatical accuracy (Bowen et al., 1985), consisted of a 40-item task requiring the insertion of "a," "an," "the," or the zero article (ø) into a text. Care was taken to ensure the test's validity and appropriateness for the cohort. The text passages were contextualized within the field of medicine to align with student backgrounds, and their readability was controlled via the Flesch Reading Ease formula (Gay & Airasian, 2000). For scoring, each of the 40 items was worth one point. A sample instrument is provided in Supplementary Material A.

Writing tasks

To complement the quantitative data from the cloze tests, we developed two writing tasks specifically to observe how students processed and applied corrective feedback. The first was a guided error correction task where participants located and rectified article mistakes in passages drawn from medical texts. This controlled design served a critical purpose: it allowed us to standardize the feedback provided to each group, creating a reliable basis for comparison (Ferris, 2003). Following this, a more open-ended sentence rewriting task was administered. This second task challenged students to move beyond simple error identification and instead reconstruct sentences, compelling them to demonstrate their grasp of article rules within a meaningful syntactic and semantic context. This provided valuable insight into their ability to transfer learning to less constrained scenarios.

Exit questionnaire

To investigate learner perceptions, an exit questionnaire was administered to all participants upon completion of the delayed post-test. The instrument consisted of 15 items scored on a 7-point Likert scale (1 = Strongly Disagree, 7 = Strongly Agree), designed to elicit detailed opinions on the AI, verbal, written, and hybrid feedback conditions. The complete instrument is available in Supplementary Material B.

The questionnaire items were developed to measure three key constructs, informed by prior research in second language feedback (Bitchener, 2008). These were: (a) Cognitive Understanding: participants' self-reported comprehension and application of the feedback; (b) Emotional Response: their affective reactions to the feedback, such as stress or motivation; and (c) Perceived Effectiveness: their appraisal of the feedback's impact on their grammatical skills. This multi-faceted design aimed to provide a comprehensive picture of student engagement.

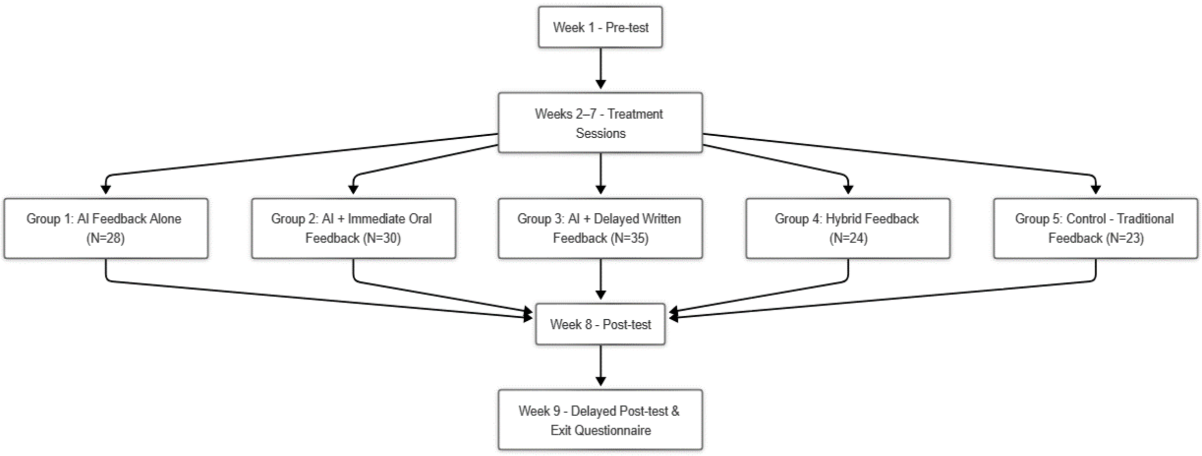
The instrument's psychometric properties were established prior to the main study. First, content validity was confirmed through an expert review process involving two applied linguistics specialists who provided recommendations for improving item clarity and relevance, which were subsequently incorporated (Creswell & Garrett, 2008). Second, reliability was assessed via a pilot test with a non-participating student group. A Cronbach's alpha analysis of the pilot data produced a coefficient of 0.85, confirming a high degree of internal consistency for the scale.

Procedure

The flowchart illustrates the sequence of activities, including the pre-test, multiple treatment sessions, feedback mechanisms, post-tests, and the final delayed post-test with an exit questionnaire. Each step indicates the progression of tasks and assessments aimed at evaluating the effects of different feedback types (AI-generated and teacher-provided) on participants' article usage in English.

Figure 1

Flowchart of the Experimental Procedure Conducted over Nine Weeks



Data Analysis

Statistical analysis was performed using SPSS. Descriptive statistics (means and standard deviations) were calculated to examine the overall improvement from pre-test to post-test and delayed post-test. Normality was assessed, ensuring the suitability of parametric tests. Mixed between-within-subjects ANOVA was conducted to examine the interaction between time and group, assessing the effectiveness of the various feedback types across three-time points (pre-test, post-test, delayed post-test). Additionally, post-hoc comparisons using Tukey's Honestly Significant Difference (HSD) test were performed to identify significant differences between groups. Independent t-tests were employed to verify the results of post-hoc comparisons, and Pearson's correlation analysis was conducted to explore the relationship between students' attitudes toward feedback and their grammatical accuracy.

Results

Descriptive Statistics

This study analyzed five groups: AI Feedback Alone, AI Feedback + Immediate Oral Teacher Feedback, AI Feedback + Delayed Written Teacher Feedback, AI Feedback + Immediate Oral + Delayed Written Teacher Feedback, and a Control group. Each group participated in three testing phases: pre-test, post-test, and delayed post-test. The descriptive statistics for the pre-test, post-test, and delayed post-test phases are provided to evaluate changes in grammatical accuracy over time.

Pre-test

The pre-test scores were analyzed to ensure no significant differences between groups before receiving feedback. The descriptive statistics for the pre-test scores are presented in Table 2.

Table 2

Descriptive Statistics for Five Groups in the Pre-Test

Group	N	Mean	SD	Min.	Max.
AI Feedback Alone	28	16.60	2.55	12	21
AI Feedback + Immediate Oral Teacher Feedback	30	16.46	2.88	10	23
AI Feedback + Delayed Written Teacher Feedback	35	16.00	2.98	9	21
AI Feedback + Immediate Oral + Delayed Written Feedback	24	16.95	3.29	11	24
Control	23	16.26	3.79	9	23

As shown, the mean scores across groups were similar, with no significant differences in pre-test performance. The group that received AI Feedback + Immediate Oral + Delayed Written Feedback had the highest mean score (M = 16.95), while the group receiving AI Feedback + Delayed Written Teacher Feedback had the lowest mean score (M = 16.00).
A one-way ANOVA was conducted to confirm that no significant differences existed between the groups at this stage.

Table 3

One-Way ANOVA for Pre-Test Scores Across Five Groups

Source	SS	df	MS	F	p-value
Between Groups	10.13	4	2.53	0.28	0.889
Within Groups	1201.42	135	8.90		
Total	1211.55	139			

SS: Sum of Squares; df: Degrees of Freedom; MS: Mean Square; F: F-statistic

Given the p-value of 0.889, which is greater than the significance level of 0.05, we can conclude that the differences between the groups were not statistically significant. This means all groups were comparable at the start of the study, providing a solid foundation for further analysis of the post-test and delayed post-test scores.

Effectiveness of AI Feedback and Teacher Feedback Combinations (Research Questions 1 & 2)

The effectiveness of each feedback type was assessed through post-test and delayed post-test scores. The following tables present descriptive statistics for the post-test and delayed post-test phases, followed by the interpretation of the key findings.

Post-test

The post-test scores demonstrated notable improvements in the correct use of articles across all experimental groups. The group receiving AI Feedback + Immediate Oral + Delayed Written Feedback (N=24) had the highest mean score (M=21.83, SD=3.67). The AI Feedback + Immediate Oral Teacher Feedback group (N=30) had a mean score of 19.40 (SD=2.45), followed by the AI Feedback Alone group (N=28) with a mean of 19.82 (SD=2.70), and the AI Feedback + Delayed Written Teacher Feedback group (N=35) with a mean of 18.03 (SD=2.92). The Control group (N=23) had the lowest mean (M=17.26, SD=4.94).

Table 4

Descriptive Statistics for Post-test

Group	N	Mean	SD	Min.	Max.
AI Feedback + Immediate Oral + Delayed Written Feedback	24	21.83	3.67	15	31
AI Feedback + Immediate Oral Teacher Feedback	30	19.40	2.45	15	24
AI Feedback Alone	28	19.82	2.70	13	25
AI Feedback + Delayed Written Teacher Feedback	35	18.03	2.92	12	23
Control	23	17.26	4.94	8	26

Delayed Post-test

The delayed post-test assessed the retention of the treatment effects. As expected, scores for most groups dropped slightly compared to the post-test, though they remained higher than pre-test scores, indicating learning retention. Once again, the AI Feedback + Immediate Oral + Delayed Written Feedback group had the highest mean score (M = 21.29, SD = 3.38). The AI Feedback + Immediate Oral Teacher Feedback group had a mean score of 19.65 (SD = 2.63), followed by AI Feedback Alone with a mean of 18.90 (SD = 2.23), and AI Feedback + Delayed Written Teacher Feedback with a mean score of 18.00 (SD = 3.43). The control group had a mean score of 17.95 (SD = 3.30).

Table 5

Descriptive Statistics for Delayed Post-Test

Group	N	Mean	SD	Min.	Max.
AI Feedback + Immediate Oral + Delayed Written Feedback	24	21.29	3.38	14	27
AI Feedback + Immediate Oral Teacher Feedback	30	19.65	2.63	15	23
AI Feedback Alone	28	18.90	2.23	16	28
AI Feedback + Delayed Written Teacher Feedback	35	18.00	3.43	10	24
Control	23	17.95	3.30	10	25

Inferential statistics

Normality test

Before conducting the inferential analysis, the normality of the data was assessed using the Shapiro-Wilk test for the pre-test, post-test, and delayed post-test scores. The results of the normality test are presented in Table 6. The p-values for all tests were above 0.05, indicating that the normality assumption was not violated. Therefore, parametric tests, such as ANOVA, are appropriate for further analysis.

Table 6

Shapiro-Wilk Test of Normality

Test Phase	Shapiro-Wilk Statistic	df	p-value
Pre-test	0.986	140	0.158
Post-test	0.983	140	0.077
Delayed Post-test	0.982	140	0.061

Effectiveness of Different Feedback Types

A repeated-measures ANOVA was conducted to evaluate the effects of the various feedback types on students' correct use of articles over time.

Table 7

Repeated-Measures ANOVA for the Effectiveness of Different Feedback Types

Source	Wilk's Lambda	F	Hypothesis df	Error df	p-value	Partial Eta Squared
Time	0.59	46.72	2	134	< 0.001	0.411
Time × Group	0.85	2.75	8	268	0.006	0.076

The results indicated a significant main effect for time (Wilk's Lambda = 0.59, $F(2, 134) = 46.72$, $p < 0.001$), which suggests that there was a statistically significant improvement in students' scores across the three test phases (pre-test, post-test, and delayed post-test). Moreover, a significant time-group interaction effect was found (Wilk's Lambda = 0.85, $F(8, 268) = 2.75$, $p = 0.006$). This indicates that the effectiveness of the feedback varied between the different experimental groups, and the degree of improvement in article usage depended on the type of feedback the students received.

Post-hoc Comparisons

Post-hoc comparisons, using Tukey's HSD test, were conducted to assess overall group differences in article usage across the post-test and delayed post-test phases, following the significant time × group interaction (Wilk's Lambda = 0.85, $F(8,268) = 2.75$, $p = 0.006$; Table 7). These comparisons aimed to identify which feedback types had the most significant impact on grammatical accuracy, with mean differences reflecting pooled treatment effects over time (Table 8).

The AI Feedback + Immediate Oral + Delayed Written Feedback group demonstrated the highest gains, significantly outperforming the Control group (mean difference = 4.57, p =

0.001) and the AI Feedback Alone group (mean difference = 3.79, p = 0.03). This hybrid feedback approach proved most effective for improving article usage. The AI Feedback + Immediate Oral Teacher Feedback group also showed significant improvement over the Control group (mean difference = 2.13, p = 0.05), though its gains were less pronounced than the hybrid group. The AI Feedback Alone group exhibited moderate improvement relative to the Control group (mean difference = 0.77, p = 0.92), while the AI Feedback + Delayed Written Teacher Feedback group showed the smallest gains among experimental groups but still outperformed the Control group (mean difference = 2.56, p = 0.03).

These results highlight the superior effectiveness of combining AI feedback with both immediate oral and delayed written teacher feedback, with immediate feedback appearing more impactful than delayed written feedback alone.

Table 8

Post-hoc Comparisons of Feedback Types in Post-Test and Delayed Post-Test Phases

Comparison	Mean Difference	p-value
AI Feedback + Immediate Oral + Delayed Written Feedback vs. Control	4.57	0.001
AI Feedback + Immediate Oral + Delayed Written Feedback vs. AI Feedback Alone	3.79	0.03
AI Feedback + Immediate Oral Teacher Feedback vs. Control	2.13	0.05
AI Feedback + Immediate Oral Teacher Feedback vs. AI Feedback Alone	1.36	0.10
AI Feedback Alone vs. Control	0.77	0.92
AI Feedback + Delayed Written Teacher Feedback vs. Control	2.56	0.03

Note: Mean differences are pooled estimates from Tukey's HSD test, reflecting average group performance differences across post-test and delayed post-test phases, following the significant time × group interaction (Wilk's Lambda = 0.85, F(8,268) = 2.75, p = 0.006, Partial Eta Squared = 0.076; Table 7). These values capture overall treatment effects rather than phase-specific raw score subtractions, consistent with the repeated-measures design.

Attitudes Toward Feedback and Their Impact on Performance

We surveyed student attitudes to address how students perceive the various feedback types and the relationship between these perceptions and performance. We correlated these attitudes with their performance on the tests.

Table 9

Multiple Comparisons Between Groups on Their Attitudes

Comparison	Mean Difference	p-value
AI + Immediate Oral + Delayed Written vs. AI + Immediate Oral	11.18	0.000
AI + Immediate Oral + Delayed Written vs. AI + Delayed Written	12.95	0.000
AI + Immediate Oral + Delayed Written vs. AI Alone	31.05	0.000
AI + Immediate Oral + Delayed Written vs. Control	32.58	0.99
AI + Immediate Oral vs. AI + Delayed Written	1.77	0.029
AI + Immediate Oral vs. AI Alone	19.87	0.99
AI + Immediate Oral vs. Control	21.40	0.000
AI + Delayed Written vs. Control	19.63	0.000

As shown in Table 9, the AI Feedback + Immediate Oral + Delayed Written Feedback group had significantly more positive attitudes compared to the other groups. This group's attitude scores were significantly higher than those of the AI Feedback + Immediate Oral Teacher Feedback group (p = 0.000), the AI Feedback + Delayed Written Teacher Feedback group (p = 0.000), and the AI Feedback Alone group (p = 0.000). However, there was no significant difference between this group and the Control group (p = 0.99).

The AI Feedback + Immediate Oral Teacher Feedback group also demonstrated significantly more positive attitudes compared to the Control group (p = 0.000) and the AI Feedback + Delayed Written Teacher Feedback group (p = 0.029). However, no significant difference was found between this group and the AI Feedback Alone group (p = 0.99).

The results show that combining AI Feedback with Immediate Oral and Delayed Written Feedback resulted in the most positive attitudes. In contrast, AI Feedback Alone and AI Feedback + Delayed Written Teacher Feedback resulted in more neutral attitudes.

Relationship Between Attitudes and Performance

The relationship between students' attitudes toward feedback and their performance on cloze tests was investigated using Pearson correlation analysis. As shown in Table 10, a positive correlation (r = 0.32, p = 0.000) was found between attitudes and performance, indicating that students with more positive attitudes toward the feedback were more likely to show improved performance on the cloze tests.

Table 10

Correlations between Students' Attitudes and Cloze Test Performance

Variable	Correlation (r)	p-value
Attitude and Performance	0.32	0.000

This positive correlation demonstrates that students who had a more favorable attitude toward the feedback they received were more likely to benefit from it in terms of improved grammatical accuracy.The average attitude scores for each group are presented in Table 11. The AI Feedback + Immediate Oral + Delayed Written Feedback group exhibited the most positive attitudes toward feedback, with an average score of 84.58, indicating that students in this group perceived the feedback as highly effective. The AI Feedback + Immediate Oral Teacher Feedback and AI Feedback + Delayed Written Teacher Feedback groups had moderately positive attitudes. In contrast, the AI Feedback Alone group and the Control group showed neutral attitudes.

Table 11
Students' Attitudes Scores

Group	Mean Attitude Score	Interpretation
AI Feedback + Immediate Oral + Delayed Written Feedback	84.58	Positive attitude
AI Feedback + Immediate Oral Teacher Feedback	73.40	Slight positive attitude
AI Feedback + Delayed Written Teacher Feedback	71.63	Slight positive attitude
AI Feedback Alone	53.53	Neutral attitude
Control	52.00	Neutral attitude

As seen in Table 11, the students in the AI Feedback + Immediate Oral + Delayed Written Feedback group had the most favorable perceptions of the feedback they received. Students in the other groups had either slightly positive or neutral attitudes.

Discussion

The present study offers clear evidence regarding the impact of different feedback conditions on the use of English articles. We found that the best improvement was achieved through a multi-layered approach, combining AI feedback with both immediate oral and delayed written teacher guidance. A key finding, however, is that standalone AI feedback was more effective than when paired only with delayed written correction. This outcome suggests that for specific grammatical rules, the timing of the feedback is crucial. It appears that delayed written input becomes less effective or even unnecessary when an immediate correction has already been provided by the AI. However, this result should not be interpreted as a universal principle; the reduced impact of delayed feedback may reflect the compressed time frame and task-specific nature of the academic writing context rather than an intrinsic weakness of delayed teacher input. In less time-sensitive instructional environments, delayed reflection could still play a critical metacognitive role (Bai & Nordin, 2025).

The superior performance of the group receiving a combination of AI, immediate oral, and delayed written feedback on both the post-test (M = 21.83) and the delayed post-test (M = 21.29) strongly affirms the value of a multi-faceted feedback approach. This finding aligns with established literature suggesting that multimodal feedback, which leverages both immediate and delayed components, optimizes cognitive processing and fosters long-term retention (Ellis, 2012; Lyster & Saito, 2010). Our results suggest a synergistic effect: immediate oral feedback likely facilitates on-the-spot correction, while delayed written comments encourage deeper cognitive reflection on grammatical rules (Karim & Nassaji, 2019; Yi, 2021). The integration of AI-driven corrections into this framework appears to amplify its effectiveness, creating a robust hybrid model for enhancing grammatical precision, as shown in personalized AI scaffolding (Adayilo et al., 2026; Woodworth, 2022). In contrast, the second key finding—that the AI-only group outperformed the group receiving AI plus delayed written feedback—presents a more complex picture that warrants closer pedagogical and cognitive examination. This apparent contradiction invites reconsideration of the assumption that “more feedback” automatically yields “better learning.” Instead, the findings suggest that excessive or poorly timed feedback streams can generate cognitive interference rather than consolidation, a nuance underexplored in previous hybrid-feedback research (Dong, 2024).

A key explanatory factor lies in the cognitive benefits of immediate feedback, a core affordance of the AI tool. Learners in the AI-only condition received corrections immediately upon task completion, allowing them to process and rectify errors while the relevant linguistic information was still active in their working memory (Liu et al., 2023). This temporal proximity between error and correction is crucial, as it enhances noticing and facilitates immediate uptake, leading to improved performance (Naseer & Khawaja, 2025; Van der Kleij & Lipnevich, 2021). Conversely, the temporal delay inherent in the AI + Delayed Written Feedback condition may have imposed a “cognitive disconnect.” By the time the teacher’s written feedback was delivered, the original context of the error was no longer salient, diminishing the feedback’s perceived relevance and overall impact (Ajjawi et al., 2022; Siregar et al., 2026). Yet, this interpretation must be balanced with findings from Crosthwaite and Sun (2025), who argued that delayed feedback can enhance retention when learners are explicitly trained in self-review strategies. The relative inefficacy of delay in the present study may therefore stem from limited metacognitive scaffolding rather than timing alone.

One plausible explanation centers on cognitive load theory and the inherent challenges of processing delayed corrective feedback. The temporal gap between committing an error and receiving feedback on it forces learners to retrospectively reconstruct the original context of the error, a mentally demanding task that consumes finite cognitive resources (Ellis, 2012). This need to recall the specific linguistic choices and their surrounding context significantly increases extraneous cognitive load, thereby hindering the learner’s capacity to process and integrate the feedback effectively (Han & Hyland, 2015). Conversely, the immediate feedback provided to the AI-only group circumvented this issue entirely. By delivering corrections while the task and the associated mental representations were still active in working memory, the AI system minimized extraneous load, allowing learners to allocate their full cognitive capacity to understanding and applying the corrections. Nevertheless, it is also possible that the AI-only group’s success reflects the repetitive, pattern-based nature of article errors. Unlike higher-order writing issues, articles lend themselves well to quick pattern recognition, meaning that the same immediacy might not generalize to more complex grammatical structures or discourse-level feedback (Huang & Mizumoto, 2025).

The superior outcomes of the AI-only group may also be attributable to factors of attentional focus and the perceived primacy of the feedback source. For this group, the AI feedback was the primary and exclusive corrective mechanism. This singular focus may have cultivated a greater dependency on, and consequently a more active engagement with, the immediate corrections provided by the system (Liu et al., 2023). In contrast, the group anticipating teacher feedback may have subconsciously relegated the AI feedback to a secondary, preliminary role. Expecting a more authoritative or nuanced review later, these students might have deferred deep processing of the initial AI corrections. This behavioral pattern—a reliance on delayed, expert feedback at the expense of immediate, automated correction—plausibly led to missed learning moments and, consequently, lower overall scores (Zarei et al., 2020). Future studies should explore this “expectation effect” by manipulating learners’ beliefs about feedback sources, as perceptions of authority may significantly modulate engagement, especially in cultures where teacher input is traditionally valued more than automated feedback (Rezai et al., 2024).

Finally, the results can be explained by the integrity of the feedback loop itself. The effectiveness of feedback hinges on its actionability, which is profoundly influenced by timing and relevance. When written feedback is substantially delayed, its practical significance can be compromised; the learner has moved on, and the opportunity for immediate, context-specific action is lost (Kartchava et al., 2020). Feedback yields the greatest benefit when it is delivered at a point of high receptivity—when the memory of the error is fresh and the learner is motivated to act upon the correction. (Zhang, 2020). The AI-only group benefited from a tight, unbroken feedback loop. The system’s immediate responses fostered continuous engagement and allowed for the swift conversion of feedback into practice, a dynamic that directly accounts for their improved learning outcomes.

Another plausible explanation for the observed results lies in the cognitive benefits of feedback predictability. The AI feedback system provided learners in the AI-only group with a consistent and standardized corrective framework. This uniformity eliminates the “noise” of instructional variation—such as differing teacher priorities or biases—that can affect human-delivered feedback (Van der Kleij & Lipnevich, 2021), thus providing a clear, unambiguous path to error correction. Students receiving this consistent input likely developed a more streamlined and coherent mental model of the target grammatical structures. Conversely, the group also receiving delayed written feedback was tasked with integrating two potentially divergent feedback streams. The need to reconcile differences between the AI’s suggestions and their teachers’ unique responses may have introduced an element of ambiguity or contradiction, ultimately hindering their ability to consolidate their learning effectively (Zarei & Rezadoust, 2020). This finding underscores the importance of alignment between AI systems and human feedback practices. Without calibration, the hybrid model risks cognitive dissonance rather than synergy—a challenge that warrants further empirical investigation (Lo et al., 2025).

Beyond performance metrics, an examination of learner perceptions provides crucial insights into the affective dimensions of feedback efficacy. The questionnaire data reveal a clear hierarchy of attitudes among the experimental groups. Notably, the cohort receiving the tripartite feedback model—AI combined with immediate oral and delayed written feedback—reported significantly more favorable perceptions than their peers. In stark contrast, both the AI-only group and the AI + Delayed Written Feedback group expressed more neutral dispositions. This divergence in learner attitudes is not a trivial finding; a positive affective response to feedback is a well-established precursor to heightened engagement, sustained motivation, and ultimately, superior learning outcomes (Andriani et al., 2026; Tsao et al., 2021). Thus, the most effective feedback ecology appears to be one that is perceived by learners as the most comprehensive and supportive.

The exceptionally favorable attitudes (M = 84.58) reported by the group receiving the three-part feedback can be interpreted through the lens of pedagogical scaffolding. This combination created a layered support structure that addressed different temporal and cognitive learning requirements. The immediate oral feedback acted as a primary scaffold, enabling learners to resolve errors and uncertainties in real time through direct interaction. Subsequently, the delayed written feedback served as a secondary, more permanent scaffold, providing a detailed artifact for deeper contemplation and the reinforcement of grammatical principles (Kartchava et al., 2020). The high level of satisfaction expressed by this group suggests a strong learner perception of this layered approach as optimally beneficial and supportive. This outcome corroborates existing research which posits that comprehensive, multi-layered feedback systems yield greater learner buy-in and effectiveness (Van der Kleij & Lipnevich, 2021).

The data reveal a notable paradox within the AI Feedback Alone group: superior performance relative to the AI + Delayed Written group coexisted with decidedly neutral attitudes (M = 53.53). This finding points to a fundamental tension between instrumental efficacy and a relational deficit in automated feedback systems. The AI tool's advantage lies in its immediacy and consistency, which evidently translated into tangible performance improvements. However, this efficacy was not mirrored in learner satisfaction, likely due to the absence of personalized, dialogic interaction. Automated feedback often cannot replicate the socio-affective support and adaptive expertise a human teacher provides (Zarei & Rezadoust, 2020). Students may therefore appraise the feedback's credibility or personal relevance with neutrality, even while benefiting from its corrective function. This outcome supports the notion that while automated feedback can be a potent tool for skill acquisition, it may not engender the same level of motivational investment or satisfaction as feedback incorporating a human element (Han & Hyland, 2015).

The attitude scores for the AI + Delayed Written Feedback group (M = 71.63) provide further evidence for the criticality of feedback timing. While these learners benefited from personalized teacher comments, their reported attitudes were markedly less enthusiastic than those of the group receiving the immediate oral component (M = 84.58). This difference suggests that the delay disrupted the formation of an effective, actionable feedback loop. For feedback to be perceived as maximally beneficial, it must be received at a point where the learner can readily connect it to the specific performance and promptly apply the advice (Ellis, 2012). Delaying the written component, despite its detail, may have severed this crucial connection. Learners, having already disengaged from the original task context, may perceive the belated feedback as an archival note rather than a dynamic tool for immediate improvement, thereby tempering their overall satisfaction with the instructional process.

The group that received AI feedback along with immediate oral teacher feedback showed a positive attitude, with an average score of 73.40. This score was significantly higher than those of both the Control group and the group that received only AI feedback. This suggests that immediate oral feedback enhances learners' perception of its value, likely because it provides opportunities for real-time clarification and interaction (Bitchener & Storch, 2016). Nevertheless, the slightly lower scores in attitudes compared to the group receiving combined feedback indicate that while immediate oral feedback is helpful, its effectiveness is even greater when accompanied by delayed written input, which allows for further review and reinforcement.

A statistically significant positive correlation was identified between learners' attitudes towards the feedback received and their subsequent performance on the cloze test ($r = .32$, $p < .001$). This finding provides robust quantitative evidence for the affective correlates of cognitive gain, suggesting that a learner's perceptual stance is not merely an incidental outcome but is intrinsically linked to their capacity for linguistic improvement. Learners who reported more favorable attitudes demonstrated greater gains in grammatical accuracy, a result that aligns with prior research positing that positive affect enhances learner uptake and engagement with corrective input (Zarei & Rezadoust, 2020). Consequently, the superior performance of the comprehensive feedback group is not an isolated phenomenon; it can be partly attributed to the highly positive motivational state fostered by the multimodal feedback design. This underscores the pedagogical imperative to design feedback ecologies that optimize not only the informational content but also the affective experience of the learner.

The outcomes for the comprehensive and AI-only feedback groups highlight the distinction between synergistic and purely functional learning. The combination of exceptional performance and highly positive attitudes in the AI + Immediate Oral + Delayed Written group is indicative of a synergistic effect, where the positive affective response and cognitive gains mutually reinforced one another. In this condition, favorable perceptions appear to act as a catalyst, transforming effective feedback into exceptional learning. Conversely, the AI-only group represents a model of functional learning. The moderate performance gains, coupled with neutral attitudes, reveal a learning experience that was effective on a technical level but motivationally and affectively sterile. This demonstrates that while an automated tool can provide correct information, it may fail to create the rich, engaging, and satisfying learning ecology necessary to maximize learner potential.

This study's findings necessitate a reconceptualization of how feedback effectiveness is evaluated, moving from a singular focus on performance to a more integrated perspective that equally values learner perceptions. True pedagogical impact, as demonstrated herein, is achieved not simply by correcting errors, but by fostering a receptive and motivated state in the learner. This points to a fundamental principle of synergy in feedback design: the whole is greater than the sum of its parts. Instructional designs that isolate feedback components—whether automated, oral, or written—are demonstrably less effective than those that weave them together into a coherent support structure. Therefore, the primary pedagogical implication of this work is an endorsement of composite feedback strategies. The exemplary success of the AI + Immediate Oral + Delayed Written feedback condition provides a robust blueprint for educators seeking to create a comprehensive learning experience that leverages the distinct yet complementary affordances of immediacy, interaction, and reflection.

Conclusion

This study investigated the differential impact of various feedback modalities on the acquisition of English article usage by Iranian EFL medical students. The results unequivocally demonstrate the superior efficacy of a synergistic feedback model that integrates AI-driven input with both immediate oral and delayed written teacher commentary. This composite approach yielded the most substantial and durable gains in grammatical accuracy, as evidenced by its leading scores on both post-test and follow-up measures. The model's success is attributable to its capacity to provide a layered support structure, facilitating both immediate cognitive uptake and subsequent reflective consolidation. Notably, the superior performance of the AI-only group over the AI plus delayed written feedback group highlights the critical role of temporal contiguity in feedback delivery. Furthermore, a significant positive correlation was established between learner attitudes and grammatical performance, affirming that affective disposition is a key mediator of learning outcomes. The comprehensive feedback condition, which also elicited the most positive attitudes, underscores a central thesis: optimal pedagogical outcomes are achieved when feedback strategies are designed to enhance both linguistic accuracy and foster positive learner perceptions, thereby creating a virtuous cycle of engagement and achievement.

This study's outcomes argue for a principled integration of educational technology, moving beyond the simple deployment of AI tools toward their thoughtful orchestration with established human feedback practices. The evidence suggests that enhanced learning is a direct function of a well-architected feedback system—one that prioritizes timeliness, promotes active learner engagement, and supports long-term consolidation. As such, this research provides actionable insights for both practitioners and scholars, charting a course for the effective co-deployment of AI and teacher-led feedback. In pedagogical contexts where grammatical accuracy is non-negotiable, the blended, multi-modal approach validated in this study represents a best-practice model for cultivating advanced linguistic competence in EFL learners.

The findings of this study yield several pedagogical implications for English as a Foreign Language (EFL) instruction, particularly in contexts such as Iranian medical programs where grammatical precision is crucial for academic and professional communication. First, the results underscore the value of integrating multimodal feedback approaches. The most effective learning outcomes were achieved when AI-generated feedback was complemented by teachers' immediate oral and delayed written input. This hybrid model capitalizes on the distinct affordances of each feedback type: AI tools provide immediacy and consistency, while teacher feedback contributes interpretive depth, clarification, and opportunities for reflection. Together, these elements create a scaffolded learning environment that promotes both immediate correction and long-term retention. A second implication relates to the timeliness and reliability of feedback. The study confirms that feedback is most beneficial when it occurs close in time to the learner's performance. AI systems offer a clear advantage in this regard, allowing learners to notice and address errors while their linguistic context remains cognitively active. However, the impact of AI feedback is amplified when paired with real-time teacher comments and later written reinforcement. Consequently, instructors should prioritize prompt and consistent feedback delivery, particularly for complex grammatical structures such as article usage, where immediacy enhances noticing and uptake. Third, the findings highlight the importance of fostering positive learner attitudes toward feedback. The correlation between favorable perceptions and improved grammatical performance suggests that affective engagement plays a key role in feedback effectiveness. Teachers can cultivate these attitudes by maintaining interactive and dialogic feedback practices—through brief in-class discussions or supportive written reflections—that help learners view feedback as constructive and collaborative rather than evaluative or punitive.

The results also caution against an overreliance on delayed written feedback. While written comments remain an important element of instruction, they proved less effective for immediate grammatical improvement when used in isolation. Delayed feedback may lose relevance once learners have moved beyond the original task. Accordingly, written feedback should serve as a complementary rather than primary tool, ideally paired with synchronous oral interaction or instant AI feedback to reinforce relevance and learner engagement. Another implication concerns learner autonomy and engagement. AI feedback was found to promote self-directed learning behaviors by offering immediate, individualized input that encouraged students to take ownership of their corrections. Teachers should therefore guide learners in interpreting and applying AI-generated feedback critically, helping them develop strategies for self-monitoring and independent revision. Such integration supports the development of self-regulated learning and reduces overdependence on teacher intervention. Finally, the study underscores the need for context-sensitive feedback practices. In EFL contexts such as medical English programs, grammatical accuracy is not merely a linguistic objective but a prerequisite for academic success. Instructors and curriculum designers should tailor feedback strategies to disciplinary and learner-specific needs, combining the efficiency of AI feedback with the pedagogical nuance of human instruction. This tailored approach can enhance both linguistic precision and learner confidence. In summary, the findings advocate for a balanced, multimodal feedback framework that unites immediacy, interaction, and reflection. When effectively combined, these feedback modalities not only strengthen grammatical accuracy but also foster positive learner attitudes and greater autonomy—cornerstones of sustainable language learning.

ORCID

https://orcid.org/0000-0003-3050-216x
https://orcid.org/0000-0002-6393-2759
https://orcid.org/0000-0001-6249-0293
https://orcid.org/0000-0003-0570-8887

Publisher's Note

The claims, arguments, and counter-arguments made in this article are exclusively those of the contributing authors. Hence, they do not necessarily represent the viewpoints of the authors' affiliated institutions, or EUROKD as the publisher, the editors and the reviewers of the article.

Acknowledgements

The authors would like to express their sincere gratitude to the participants who took part in this study and to the instructors who supported the implementation of the research procedures. We also acknowledge the support of Central Asian University for providing ethical approval and facilitating the research process.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

CRediT Authorship Contribution Statement

Afshin Soori: Conceptualization, Methodology, Investigation, Formal Analysis, Writing – Original Draft
Laleh Khojasteh: Data curation, Methodology, Supervision, Writing – Review & Editing, Project Administration
Sedigheh Shakib Kotamjani: Writing – Review & Editing, supervision- Funding Acquisition

Sanaz Mojarrad: Data Curation, Investigation, Visualization, Writing

Generative AI Use Disclosure Statement

Generative AI tools (Grok-4 by xAI) were used solely for basic text editing, minor copyediting, and formatting assistance consistent with standard word processing functions (e.g., Microsoft Word grammar/style checks). No AI was involved in generating content, analyzing data, creating code, drafting text, refining research design, or producing figures/tables/images. All content reflects the original authors' work.

Ethics Declarations

World Medical Association (WMA) Declaration of Helsinki–Ethical Principles for Medical Research Involving Human Participants

This study was conducted in accordance with the ethical standards outlined in the World Medical Association Declaration of Helsinki (2013). Ethical approval was obtained from the Institutional Review Board (IRB) of Central Asian University (Ethical Review Code: CAU-IRB-2025-032). All participants were informed about the purpose of the study, and informed consent was obtained before participation. Participation was voluntary, and all data were anonymized to ensure confidentiality. The study involved no physical or psychological risk to participants.

Competing Interests

The authors declare that there are no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data Availability

The datasets generated and/or analyzed during the current study are not publicly available due to ethical and privacy considerations but are available from the corresponding author on reasonable request. All data have been anonymized to protect participants' confidentiality.

References

Adayilo, D. M., Oyefolahan, I. O., Ndunagu, J. N., Otuya, C., Malcalm, E., & Twabu, K. (2026). AI-powered tutoring systems for personalised learning feedback in developing secondary education contexts. *Journal of Future Artificial Intelligence and Technologies*, *2*(4), 549-564. https:/doi.org/10.62411/faith.3048-3719-290

Ajjawi, R., Kent, F., Broadbent, J., Tai, J. H.-M., Bearman, M., & Boud, D. (2022). Feedback that works: A realist review of feedback interventions for written tasks. *Studies in Higher Education*, *47*(7), 1343-1356. https://doi.org/10.1080/03075079.2021.1894115

Andriani, S. S., Safitri, W., & Baluqiah, K. (2026). Students' trust in AI feedback and its effect on critical reading in EFL. *DIAJAR: Jurnal Pendidikan dan Pembelajaran*, *5*(1), 353-360. https://doi.org/10.54259/diajar.v5i1.7051

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

Banihashem, S. K., Noroozi, O., Khosravi, H., Schunn, C. D., & Drachsler, H. (2025). Pedagogical framework for hybrid intelligent feedback. *Innovations in Education and Teaching International*, 1-17. https://doi.org/10.1080/14703297.2025.2499174

Bitchener, J. (2008). Evidence in support of written corrective feedback. *Journal of Second Language Writing*, *17*(2), 102-118. https://doi.org/10.1016/j.jslw.2007.11.004
Bitchener, J., & Storch, N. (2016). *Written corrective feedback for L2 development* (Vol. 96). Multilingual Matters. http://www.jstor.org/stable/43267894
Bowen, J. D., Madsen, H. S., & Hilferty, A. (1985). *TESOL techniques and procedures*. Newbury House Publishers.
Creswell, J. W., & Garrett, A. L. (2008). The "movement" of mixed methods research and the role of educators. *South African Journal of Education*, *28*(3), 321-333. https://doi.org/10.15700/saje.v28n3a176
Crosthwaite, P., & Sun, S. (2026). Generative AI and L2 written feedback studies: A scoping review. *RELC Journal,* 57(*1*), 207–219. https://doi.org/10.1177/00336882251386530
Denny, P., Khosravi, H., Hellas, A., Leinonen, J., & Sarsa, S. (2023). Can we trust AI-generated educational content? Comparative analysis of human and AI-generated learning resources. *arXi. https://doi.org/10.48550/arXiv.2306.10509*
Dong, L. (2024). ChatGPT in language writing education: Reflections and a research agenda for a ChatGPT feedback engagement framework. *Language Teaching Research Quarterly*, *43*, 121-131. https://doi.org/10.32038/ltrq.2024.43.07
Ellis, N. C. (2012). What can we count in language, and what counts in language acquisition, cognition, and use. *Frequency Effects in Language Learning and Processing*, *1*, 7-34. https://doi.org/10.1515/9783110274059.7
Ferris, D. R. (2003). Response to student writing: Implications for second language students. *Lawrence Earlbaum Associates*.
Gay, L. R., & Airasian, P. W. (2000). *Student guide to accompany educational research: Competencies for analysis and application*. Merill.
Han, Y., & Hyland, F. (2015). Exploring learner engagement with written corrective feedback in a Chinese tertiary EFL classroom. *Journal of Second Language Writing*, *30*, 31-44. https://doi.org/10.1016/j.jslw.2015.08.002
Huang, J., & Mizumoto, A. (2025). The role of generative AI in mediating L2MSS and engagement with written feedback in EFL learning: A structural equation modeling approach. *Annual Review of Applied Linguistics*, 1-17. https://doi.org/10.1017/S0267190525000029
Hwang, G.-J., & Chien, S.-Y. (2022). Definition, roles, and potential research issues of the metaverse in education: An artificial intelligence perspective. *Computers and Education: Artificial Intelligence*, *3*, Article 100082. https://doi.org/10.1016/j.caeai.2022.100082
Jegede, O. O. (2024). Artificial Intelligence and English Language Learning: Exploring the Roles of AI-Driven Tools in Personalizing Learning and Providing Instant Feedback. *Universal Library of Languages and Literatures*, *1*(2). https://doi.org/10.70315/uloap.ullli.2024.0102002
Kaliisa, R., Misiejuk, K., López-Pernas, S., & Saqr, M. (2026). How does artificial intelligence compare to human feedback? A meta-analysis of performance, feedback perception, and learning dispositions. *Educational Psychology*, *46*(1), 80-111. https://doi.org/10.1080/01443410.2025.2553639
Karim, K., & Nassaji, H. (2019). The effects of written corrective feedback: A critical synthesis of past and present research. *Instructed Second Language Acquisition*, *3*(1). https://doi.org/10.1558/isla.37949
Kartchava, E., Gatbonton, E., Ammar, A., & Trofimovich, P. (2020). Oral corrective feedback: Pre-service English as a second language teachers' beliefs and practices. *Language Teaching Research*, *24*(2), 220-249. https://doi.org/10.1177/1362168818787546
Liang, K.-H., Davidson, S., Yuan, X., Panditharatne, S., Chen, C.-Y., Shea, R., Pham, D., Tan, Y., Voss, E., & Fryer, L. (2023). ChatBack: Investigating methods of providing grammatical error feedback in a GUI-based language learning chatbot. In *Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications* (BEA 2023) (pp. 83–99). Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.bea-1.7
Liu, C., Hou, J., Tu, Y.-F., Wang, Y., & Hwang, G.-J. (2023). Incorporating a reflective thinking promoting mechanism into artificial intelligence-supported English writing environments. *Interactive Learning Environments*, *31*(9), 5614-5632. https://doi.org/10.1080/10494820.2021.2018616
Lo, N., Chan, S., & Wong, A. (2025). Evaluating teacher, AI, and hybrid feedback in English language learning: Impact on student motivation, quality, and performance in Hong Kong. *Sage Open*, *15*(3), Article 21582440251352907. https://doi.org/10.1177/2158244025135290
Lyster, R., & Saito, K. (2010). Oral feedback in classroom SLA: A meta-analysis. *Studies in Second Language Acquisition*, *32*(2), 265-302. https://doi.org/10.1017/S0272263109990520

Namaziandost, E., & Rezai, A. (2024). Special issue: Artificial intelligence in open and distributed learning: Does it facilitate or hinder teaching and learning? International *Review of Research in Open and Distributed Learning*, 25(3), i–vii. https://doi.org/10.19173/irrodl.v25i3.8070

Naseer, F., & Khawaja, S. (2025). Mitigating conceptual learning gaps in mixed-ability classrooms: A learning analytics-based evaluation of AI-driven adaptive feedback for struggling learners. *Applied Sciences*, *15*(8), 4473. https://doi.org/10.3390/app15084473

Nasiri, E., & Shokrpour, N. (2024). Identification of factors influencing professional identity development in medical students at basic sciences stage. *Journal of Advances in Medical Education & Professionalism*, *12*(1), 45. https://doi: 10.30476/JAMP.2023.99829.1856.

Nassaji, H., & Kartchava, E. (2020). *Corrective feedback and good language teachers*. In C. Griffiths & Z. Tajeddin (Eds.), Lessons from good language teachers (pp. 151–166). Cambridge University Press.

Otaki, B. (2023). *Feedback in the era of generative AI: An explorative focus group study of student and educator perspectives in higher education* (Master's thesis, University of Gothenburg, Department of Education, Communication & Learning). GUPEA.

Rezai, A., Namaziandost, E., & Hwang, G.-J. (2024). How can ChatGPT open promising avenues for L2 development? A phenomenological study involving EFL university students in Iran. *Computers in Human Behavior Reports*, *16*, Article 100510. https://doi.org/10.1016/j.chbr.2024.100510

Shen, R., & Chong, S. W. (2023). Learner engagement with written corrective feedback in ESL and EFL contexts: a qualitative research synthesis using a perception-based framework. *Assessment & Evaluation in Higher Education*, *48*(3), 276-290. https://doi.org/10.1080/02602938.2022.2072468

Siregar, F. R., Harida, E. S., & Najiah, N. (2026). Reducing morphosyntactic errors of students' English writing through AIpowered language learning tools. *English Learning Innovation*, *7*(1), 48-65. https://doi.org/10.22219/englie.v7i1.43264

Tsao, J.-J., Tseng, W.-T., Hsiao, T.-Y., Wang, C., & Gao, A. X. (2021). Toward a motivation-regulated learner engagement WCF model of L2 writing performance. *Sage Open*, *11*(2), Article 21582440211023172. https://doi.org/10.1177/21582440211023172

Van der Kleij, F. M., & Lipnevich, A. A. (2021). Student perceptions of assessment feedback: A critical scoping review and call for research. *Educational Assessment, Evaluation and Accountability*, *33*, 345-373. https://doi.org/10.1007/s11092-020-09331-x

Wei-Xun, L., & Jia-Ying, Z. (2024). Impact of AI-driven language learning apps on vocabulary acquisition among English learners. *Research Studies in English Language Teaching and Learning*, *2*(1), 1-11. https://doi.org/10.62583/rseltl.v2i1.32

Woodworth, J. (2022). *Hybrid feedback: The efficacy of combining automated and teacher feedback for second language academic writing development* (Doctoral dissertation, York University, Toronto, Canada).

Yi, D. (2021). The effects of immediate versus delayed teacher feedback on L2 writing. *Studies in English Language Teaching*, *9*(3), 69-78. https://doi.org/10.22158/selt.v9n3p69

Zarei, A. A., & Rezadoust, H. (2020). The effects of scaffolded and unscaffolded feedback on EFL learners' speaking anxiety and speaking self-efficacy. *Journal of Modern Research in English Language Studies*, *7*(4), 111-132. https://doi.org/10.30479/JMRELS.2020.13464.1655

Zarei, M., Ahour, T., & Seifoori, Z. (2020). Impacts of implicit, explicit, and emergent feedback strategies on EFL learners' motivation, attitude and perception. *Cogent Education*, *7*(1), Article 1727130. https://doi.org/10.1080/2331186X.2020.1727130

Zhang, Z. V. (2020). Engaging with automated writing evaluation (AWE) feedback on L2 writing: Student perceptions and revisions. *Assessing Writing*, *43*, Article 100439. https://doi.org/10.1016/j.asw.2019.100439

Zheng, C., Wang, K., Gao, S., Yu, Y., Wang, Z., & Tang, Y. (2024). Design of multi-modal feedback channel of human–robot cognitive interface for teleoperation in manufacturing. *Journal of Intelligent Manufacturing*, 36, 4283–4303. https://doi.org/10.1007/s10845-024-02451-x